Doctoral School in Cognitive and Brain Science

XXX cycle

Ph.D. dissertation

Constructing multiple objects across space and time

Evelyn Muschter

Supervisor: David P. Melcher

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

For us to interact with our environment we must have an understanding of the location and properties of the objects within it over time. Incoming stimuli from different sensory modalities continuously provide information. In order to create a coherent percept, it is necessary to combine these sensory signals in the brain into a single representation of an object or event. Temporal integration, the process of combining information over time is one important prerequisite for it. Likewise, multisensory integration, the process of combining information from different sensory modalities to generate a coherent percept is of great importance. Furthermore, as our environment is not static but dynamic it is necessary to combine sensory information over time and space. Despite, the ever-increasing neuroscientific results, much about the underlying brain mechanisms remains unknown. This thesis describes an effort of understanding how sensory information are integrated in the brain and what effects attention has on how we integrate multiple objects in space and time.

Chapter 1 introduces the reader to the existing literature on neural oscillations, specifically alpha band frequency, which is hypothesized to play a critical role in temporal processing and attention. This is followed by a review of the literature on temporal integration and the theoretical background of multisensory integration. The following chapters will report the three studies conducted: Chapter 2 describes the first study, a MEG study that investigates the role of alpha oscillations in temporal integration. Chapter 3 reports the second MEG study, which focuses the role of attention to modality in multisensory integration. Chapter 4 presents the third study, a behavioral oscillations study, which focuses on the role of neural oscillations in selective attention in a dynamic scene. Finally, Chapter 5 summarizes all the findings from this research.

CHAPTER 1: OVERVIEW AND GENERAL INTRODUCTION

We live in a world where sensory input arrives continuously over time from all the different senses. This input is merged in the brain, resulting in a seamless perception of our environment. The brain must parse this flow of information into coherent objects and events. This involves combining information for the same event or object over time and across senses. The overall goal of this thesis is to shed light on how the human perceptual system constructs multisensory objects in space and time. In particular, we investigated the role of neural oscillations on the construction of spatio-temporal percepts. To this end, we conducted three experiments. The thesis is structured the following way:

In Chapter 1, I will begin by introducing neural oscillations. In particular, I will focus on the alpha frequency band, which is hypothesized to play a critical role in attention and in temporal processing of stimuli into a coherent perceptual event in terms of objects over time. Then, I introduce the concept of temporal integration windows as a way to characterize whether stimuli are combined over time or treated as unique events, and the potential link between temporal integration windows and oscillations. In the final part of the chapter, I will focus on multisensory integration and the way input is combined across senses.

In Chapter 2, I describe an MEG experiment that investigated the role of neural oscillations in the temporal integration and segregation of stimuli in rapid succession. We tested the hypothesis that pre-stimulus alpha oscillations influence perceptual outcome. To this end, we utilized a paired-stimulus paradigm with a variable temporal gap between the presentation of a near-threshold flash and an above-threshold flash while recording MEG. Evidence for alpha band frequency as a predictor of perceptual outcome is presented and discussed.

Chapter 3 focuses on the spatial-temporal integration of information across the senses. In a MEG study, we addressed the question of how attention to sensory modalities, within a multisensory

stimulus, would alter pre-stimulus neural oscillation and the way in which the stimulus was processed in sensory cortices. The main findings were that attention to modality is reflected in pre-stimulus alpha oscillations as well as in early evoked responses in the constituent sensory regions.

In Chapter 4 we further investigated the role of neural oscillations in selective attention in a dynamic scene. In a behavioral experiment we investigated two different multiple object tracking strategies and measured fluctuations in accuracy over time ("behavioral oscillations": Landau & Fries, 2012). The main finding was that performance fluctuated rhythmically in both tasks, although at different frequencies.

Finally, Chapter 5 reviews and summarizes the main findings.

1.1. Neural oscillations and their role in sensory processing

Neural oscillations, also described as brain waves, are the rhythmic pattern of neuronal activity in the central nervous system. Brain waves are recordable via electroencephalography (EEG) and magnetoencephalography (MEG). These measurement devices are ideal as they provide high temporal resolution, which is necessary in order to reflect the synchronous waxing and waning of summed postsynaptic activity of large neuronal populations (Wang, 2010). In other words, neural oscillations reflect the rhythmically fluctuating excitability states of neuronal populations across different temporal and spatial scales. Signals recorded from EEG and MEG have thus proven to be powerful tools in the attempt of understanding brain functions as they provide a multidimensional resolution comprised of time, space, frequency, as well as power and phase of a given frequency band. However, oscillatory pattern can also be measured by means of behavioral outcomes. Meaning, the dynamics of a given task can be mapped out by repeatedly measuring reaction times and/or accuracy to an event at equally spaced out, different points in time, in order to create time series – sets of collective ordered observations of quantitative characteristics (Kendall and Buckland, 1971).

Perhaps the most fundamental and extensively studied parameter of neural oscillations is frequency, which is defined as the number of cycles per second and expressed in Hertz (Hz). Research has established different frequency bands, which are in turn associated with different functional brain states. Although the exact boundaries are rather loosely defined and often a subject of scientific debate, they are divided into delta (0-4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-30 Hz), low gamma (30-60 Hz), high gamma (60-90 Hz) and high frequency oscillations (above 100 Hz; VanRullen and Dubois, 2011). Interactions between different neural populations are reflected in different frequency band oscillations. For example, Buschman et al. (2007) showed that while gamma band synchronization between prefrontal and parietal areas of the monkey brain were related to bottom-up processes, top-down processes were reflected in beta band synchronizations. Findings as such provide support for the idea that information is transmitted through different frequency bands, depending on their directionality (top-down/bottom-up). Furthermore, a series of reports found that in unisensory cortical areas bottom-up processing is reflected in gamma-band (30 Hz and above) frequencies, whereas top-down processes are predominantly reflected in lower frequency bands (below 30 Hz) (Fell et al., 2003; Fontolan et al., 2014; Fries, 2015; Frey et al., 2015; VanRullen, 2016).

Power is another informative parameter of neural oscillations and is expressed as the squared amplitude of the energy at a given time-frequency point. Variations in power result from a change in synchronization of the underlying neuronal population. Furthermore, power is inversely related to its frequency. In other words, signals with high frequency usually exhibit a lower power compared to low frequency signals (Buzsáki and Draguhn, 2004).

Phase is the parameter of neural oscillations that defines the position of a signal at a specific time point along the oscillatory wave and ranges from 0-360°. Phase is an informative measure of the timing of neuronal activity and thus has been suggested to define discrete windows of excitation and inhibition with high temporal resolution (Busch et al., 2009; VanRullen, 2016b).

Neuronal oscillations can generally be categorized in induced and evoked responses. On one hand, evoked responses stem from an onset of an external event, such as the onset of a stimulus. They are commonly expressed as event-related potentials (ERPs) or event-related fields (ERFs), the summation over trials of identical phase. On the other hand, induced oscillations can occur independent of external stimulation. Therefore, they might not be time-and phase-locked to the onset of a stimulus. Instead, cognitive processes such as attention can modulate induced oscillations. Taken together, data from these studies are supportive of a model that divides interactions between neuronal populations into encoding (reflected in gamma-band oscillations) and integrative (reflected in lower frequency oscillations) (Gratton, 2018). Such model suggests that the brain segregates information according to their directionality into different frequency channels (Donner and Siegel, 2011; Gratton, 2018). Moreover, empirical findings from these studies complement the hypothesis that neural oscillations enable information transfer in an efficient manner among different brain areas through specific and dynamic neuronal networks (Fries, 2009).

1.2. Evidence for a specific role of alpha band oscillations in perception

One of the most extensively studied frequencies is the alpha band. Alpha is commonly defined as the frequency band between 8 and 12 Hz (Lange et al., 2014). Alpha frequency is the most prominent rhythm in the human cortex, and is often already visible in raw EEG and MEG recordings. When Berger first scientifically described alpha, it was thought to be reflective of states of low arousal because an occipital alpha amplitude increase resulted from subjects closing their eyes (Berger, 1929; Adrian and Matthews, 1934). For years, alpha band power was interpreted as a neural correlate of cortical idling. Many scientific investigations in the last decades have shown that alpha is not just merely a passive brain rhythm.

Studies in attention have shown that active top-down control (task/ goal directed) is reflected in alpha band activity (for a review see Frey et al., 2015). Intersensory and multisensory experiments

have shown that alpha power increases in task-irrelevant brain areas prior to stimulus onset when attention is directed towards a different sensory modality (Fu et al., 2001; Bauer et al., 2012; Mazaheri et al., 2013). For example, Mazaheri et al. (2013) investigated pre-stimulus alpha and beta power modulations when participants attended to one modality, while suppressing another distracting sensory modality. When participants attended to the visual modality (judging the orientation of different gratings), while ignoring a temporally synchronous auditory distractor (tones of different pitches), alpha power in the early visual cortex decreased along with a 14-16 Hz increase in the right supramarginal gyrus, a region that regulates auditory pitch processing. Furthermore, trial-by-trial pre-stimulus alpha modulations in these task-specific cortical regions were correlated with the respective task discrimination times only when the target was presented with a distractor, suggesting that region-specific alpha power modulations have functional consequences for stimulus processing.

Likewise, spatial attention, which describes directing attention to a specific location, is also reflected in alpha band power in corresponding brain regions (Thut et al., 2006; Busch and VanRullen, 2010; Jones et al., 2010). Data from EEG and MEG studies showed that pre-stimulus alpha band power in sensors contralateral to the attended spatial location over the corresponding sensory brain areas is decreased, while alpha band power ipsilateral to the attended spatial location exhibits an increase from baseline (Jones et al., 2010). Alpha power modulations were found in the pre-stimulus time period when attention was directed towards the location of an upcoming visual stimulus contralateral to the cued location in primary visual areas. Thut and colleagues (2005) recorded EEG during a spatially cued target detection task. Subjects were instructed to attend to a cued location (right or left) and respond if the presented visual stimuli appeared in the cued or un-cued location. The researchers found contralateral alpha (8-14 Hz) modulations that were correlated with subject's response times. Van Ede and colleagues (2011) utilized a tactile discrimination task in which participants' attention was cued with full validity to either the right or left hand, and found spatially specific contralateral suppression, and ipsilateral increase effects of alpha and beta power in

the primary sensory cortex. This effect was more pronounced for the beta band (15-30 Hz) and was highly consistent across subjects. Importantly, contralateral beta power decreased with time towards the anticipated stimulus onset, which shows that somatosensory beta power modulations are produced by contralateral suppression. Interestingly, because the tactile stimulation was to occur either at 1, 2 or 3 s, after the cue, they were also able to investigate if these power modulations exhibited temporal specificity. They found that lateralized alpha and beta power modulations track the temporal position at which an expected event is to occur. Thus, their results indicate that not only spatial attention, but also temporal specificity is reflected in pre-stimulus modulations of alpha and beta oscillations in the primary somatosensory cortex. Müller and Weisz (2011) found an asymmetric modulation of auditory alpha in the right auditory cortex as pre-stimulus alpha power increased when participants' attention was directed to the right. This is plausible due to the functional specialization of the auditory cortex: while left hemisphere of the auditory cortex is specialized for rapid temporal processing, the right hemisphere of the auditory cortex is specialized for spectral processing (Zatorre and Belin, 2001). They additionally found that the right auditory cortex was functionally connected with the frontal eye-field during an ipsilateral alpha power increase (Müller and Weisz, 2011). Thus, the alpha band rhythm reflects an active, spatially specific top-down influence of sensory processing.

Importantly, alpha band oscillations are not just reflective of attentional control, but also reflect conscious perception of visual, auditory and tactile stimuli (Ruhnau et al., 2014). Conscious perception of visual stimuli is negatively correlated with alpha power over posterior EEG/MEG sensors corresponding to the visual cortex (Ergenoglu et al., 2004; Van Dijk et al., 2008; Wyart and Tallon-Baudry, 2009; Iemi et al., 2017). Furthermore, neuromodulation studies have demonstrated that cortical excitability fluctuates along with ongoing alpha rhythm in a series of transcranial magnetic stimulation studies by Romei and colleagues (Romei et al., 2008a; 2008b; 2010). They investigated participants' individual stimulation threshold of phosphene (visual illusory percepts) perception. This threshold was positively correlated with individual resting state alpha power (Romei

et al., 2008b). An active entrainment of the alpha rhythm in the visual cortex impaired participants' ability to detect near-threshold stimuli in the hemifield contralateral to the stimulated side (Romei et al., 2010). By means of MEG recordings, Leske et al. showed pre-stimulus alpha power decreases in the primary auditory cortex prior to the successful detection of an auditory near-threshold stimulus (Leske et al., 2015). In addition, similar alpha power findings have been reported in the somatosensory domain (Schubert et al., 2009; Weisz et al., 2014).

Finally, the phase of the ongoing alpha oscillations regulates the temporal aspects of input processing (Busch et al., 2009). According to the inhibition-timing hypothesis, stimulus processing is facilitated during troughs, and hindered during peaks of the ongoing oscillatory activity (Mathewson et al., 2009; Klimesch et al., 2007). Furthermore, the phase of oscillatory alpha activity can be entrained with rhythmically presented stimuli prior to the actual target stimuli presentation (Ronconi and Melcher, 2017) or reset by stimuli from another sensory modality (Romei et al., 2012). It must be pointed out that in accord with the function inhibition hypothesis discussed earlier: both alpha phase dependency and alpha band entrainment are contingent on high pre-stimulus alpha power (Jensen and Mazaheri, 2010).

Taken together, alpha band activity is linked to cortical activity and takes on a modulatory role for sensory perception. Data from neuroscientific studies, such as the ones discussed here, have facilitated researchers to form the specific hypothesis of gating through inhibition (Jensen and Mazaheri, 2010). Alpha band activity is a sensory gating mechanism that regulates information processing between sensory and higher-order cortices. In summary, there is much scientific agreement that high alpha inhibits task-irrelevant areas and connections, while low alpha enables efficient information processing in task-relevant areas.

2.1. Temporal integration windows: combining or segmenting sensory input over time

Although sensory input is continuously arriving over time, it needs to be parsed by our perceptual system into specific objects by combining input over time. In the following paragraph, I will discuss vision as a continuous process through time. Visual perception is often modeled as a process of chunked instantaneous visual input in forms of 'snapshots' or frames. These 'snapshots' can integrate information over considerable periods of time, identified as temporal integration windows. Due to the diverse nature of the stimuli our visual system encounters, different temporal integration windows have been proposed (Melcher et al., 2014). Durations shorter than 30-40 ms cannot be discriminated by the visual system (Mach, 1965) and are instead experienced as a time point rather than a continuance. While around 40 ms is commonly considered the lower threshold, data from temporal integration studies has facilitated researchers to accumulate evidence for temporal integration windows that extends to 2-3 seconds (Hasson et al., 2008; Fairhall et al., 2014). Various masking and integration studies have tested the concept of time windows of around 200-300 ms (Jiang et al., 2005; Fracasso et al., 2010; Melcher and Fracasso, 2012) as well as around 100 ms (Enns and Di Lollo, 2000; Breitmeyer and Öğmen, 2006; Wutz et al., 2014). When stimuli are presented in rapid succession at the same location, the temporal offset, their interstimulus interval (ISI), is the determining factor of perceptual outcome. Following the concept of temporal windows: short ISIs of around 50 ms support integration, where the two stimuli are temporally and spatially fused together and perceived as one. An increase in ISI annihilates integration in favor of segregation and the stimuli are perceived in its true quantity. Data from a variety of studies adduce evidence in support of such temporal integration windows (VanRullen and Koch, 2003; van Wassenhove et al., 2007; Cecere et al., 2015; Samaha and Postle, 2015; Baumgarten et al., 2017).

For example, with a variation of the missing element task (Di Lollo, 1980) we previously showed that temporal integration and segregation for visual stimuli can be seen as complementary tasks (Wutz et al., 2016). Figure 1 shows our experimental set-up as follows: two successive visual

displays of seven full and one half circle each at random locations in a symmetrical lattice with a variable ISI in between are presented. Importantly, the individual element locations of the two displays did not overlap; except for in the case of the half circles, where the two circle halves were in the same complimentary location. Hence, a complete grid of 16 elements was defined over the two stimulus frames and ultimately left one location empty. In integrations trials, participants were to indicate the location where no circle was presented in both displays. While in segregation trials, participants were asked to identify the location in which the half circles were presented.

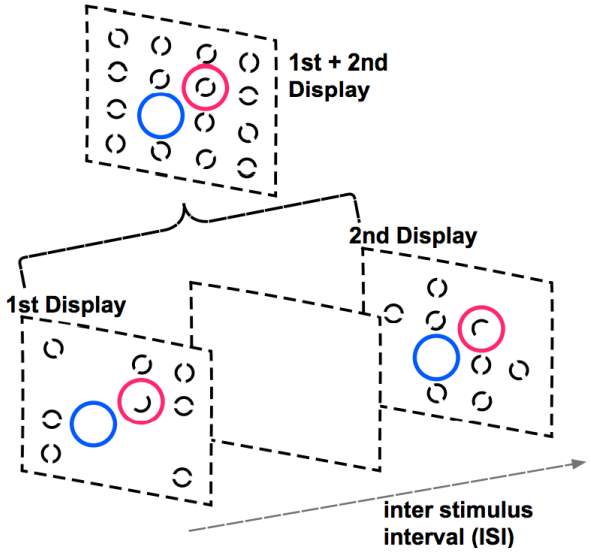

Figure 1. Experimental setup for integration/segregation task in Wutz et al., 2016. The superimposed blue circles specifies the integration task, while the superimposed red circles indicates the segregation task. The combination of display 1 and 2 highlight both tasks combined within the same stimuli.

This set up enabled us to investigate the temporal dynamics (by means of varying the ISI) of visual integration versus segregation by changing the task, while keeping the stimulation the same. While integration performance declined with longer ISI, segregation performance increased. Average task performance was equivalent at 68 ± 27 ms ISI for both tasks (Wutz et al., 2016). These results are in support of the idea that integration windows serve a function of preserving visual stability and help coordinate further visual processing. Integration of successive stimuli can be interpreted as the

brain's way of coping with the processing lag of sensory input by accumulating visual information over longer meaningful intervals before processing and hence reducing perceptual effort. This consequently results in a loss of temporal resolution within the period of integration. However, given a sufficient latency, changes in the environment are still rapidly detectable by perceiving sequential stimuli as disjoint. Segregation can be interpreted as a form of habituation, a process to avoid error of integration through a reduction of blending and enhancement of change detection (Huber, 2014). If a sequence of stimuli is presented at the right time, the percept accurately reflects the sequence. It can be attributed to the fact that the first stimulus has become habituated and is no longer apparent by the time the second stimulus appears (Huber, 2014).

Conclusively, temporal integration windows can be seen as perceptual cycles in which the brain integrates input over a certain window length or cycle (Baumgarten et al., 2017). Now when successive stimuli fall within a certain time interval they are perceptually integrated. On the contrary, if stimuli fall in two temporal windows they are perceived as distinct events.

2.2. Underlying brain oscillations of temporal integration

Temporal integration windows provide an intrinsic concept in the quest of understanding perception. To date, a few studies provide direct evidence for potential underlying neuronal mechanisms. The most compelling candidate being neural oscillations (VanRullen et al., 2014; Samaha and Postle, 2015; Wutz et al., 2016). Several behavioral studies have shown that perception and behavior exhibit cyclic or rhythmical pattern (Landau and Fries, 2012; Drewes et al., 2015; Wutz et al., 2016; Ronconi and Melcher, 2017). Such findings have further been supported by studies using EEG/ MEG. Cycles of specific neuronal oscillations form the potential mechanism for temporal integration windows and furthermore correlate with behavioral outcomes.

Data from EEG/MEG studies facilitated researchers to show that the phase of an ongoing neural oscillation is linked to perception (Busch et al., 2009; Drewes and VanRullen, 2011;

VanRullen et al., 2011; Ronconi and Marotti, 2017; Ronconi et al., 2017). We (Wutz et al., 2016) showed this for example, in the previous described integration/segregation study in both MEG and behavioral measures for both tasks with opposite phase in the theta frequency (3-5 Hz).

Recently further work from our lab (Ronconi et al., 2017) has used multivariate decoding of EEG data to show that temporal integration/segregation was dependent on the pre-stimulus phase of parieto-occipital regions. In order to investigate if different temporal integration window lengths were reflected in different frequencies, they used two different integration tasks: A two-flash fusion task, where two flashes, separated by a brief temporal gap (variable ISI) appear in the same location and an apparent motion task, where such two flashes appear at different locations. They found the highest decoding accuracy for the two-flash fusion task in the phase of alpha, but for the apparent motion in the phase of theta. The phase of alpha oscillations seems to temporally organize incoming input and prevent information overload (Gips et al., 2016).

VanRullen's wagon wheel illusion (VanRullen et al., 2006) provides evidence for a correlation between perception and alpha band power in occipital areas. In this paradigm, a constantly rotating wagon wheel sometimes creates the illusion of reversing the rotational direction. The researchers found a 13 Hz power decrease just before the onset of the illusionary percept as well as an increase just before participants transitioned back to perceiving the real motion direction. These findings offer support to the idea that motion is perceived through short temporal windows (< 100 ms).

With their publication, Cecere et al. (2015) further provided causal evidence for alpha oscillations setting temporal integration windows. They utilized transcranial alternating current stimulation (tACS) combined with EEG in a double-flash illusion task. The illusionary percept of two visual flashes arises when a brief flash is presented with two auditory stimuli at short latencies (Sham et al., 2000). The researchers showed that participants' individual occipital-parietal peak alpha frequency was correlated with the illusionary percept time window. They then utilized tACS to

manipulate that peak frequency and to that end the length of the individual alpha cycles, which modulated the illusory temporal window (Cecere et al., 2015).

2.3. Individual alpha peak frequency

Much research has investigated individual alpha peak frequency (IAPF) and its relationship to cognition (for a review see Klimesch, 1999; 2012). Alpha frequency shows a lot of inter- as well as intra-individual variations but is generally defined within the ranges of 8-14 Hz. With his publication, Klimesch (1999) set the definition of IAPF as the maximum power value in the EEG frequency spectrum between 7.5 and 12.5 Hz of an individual. Findings from a variety of studies have shown that individuals with higher IAPF show better cognitive performance (Klimesch et al., 1993; Clark et al., 2004; Jin et al., 2006; Grandy et. al, 2013a). Per contra, a variety of neurological pathologies exhibit low IAPF such as Alzheimer's disease (Moretti et al., 2004), Attention deficit hyperactivity disorder (Lansbergen et al., 2011) and traumatic brain injury (Dockree et al., 2004). Furthermore, IAPF is a highly stable measure in healthy adults, a neurophysiological trait marker that is not easily modifiable (Kondacs and Szabo, 1999; Grandy et al., 2013b; Campisi and La Rocca, 2014). As alpha oscillation has been proposed as the gatekeeper of temporal integration windows via the timing of functional inhibition, faster IAF allow for "faster alternations between 'open' and 'closed' states of information transfer" (Grandy et al., 2013a) and thus orchestrate an effective integration of information.

3.1. Multisensory Integration: combining information about the same object or event across the senses

The brain does not just have to make sense of various incoming visual stimuli. In our daily life, we encounter a plethora of different sensory stimulation besides visual information. Imagine you are in the middle of Time Square in New York City. For the first time in your life, you are in a big city and

you have never experienced such spectacle. Your senses are consistently bombarded with a plentitude of different information: buzzing sounds, flashing lights and different scents. The brain tries to make sense of such noisy data by integrating inputs coming from different sensory modalities when these are presented in close temporal or spatial proximity, appearing to come from a common source. For example, you will react to approaching dangers much faster when they are perceived through multiple senses. Hence, when you see and hear a rowdy taxi driver approaching while you are trying to cross the street, you hopefully will avoid collision as both sensory inputs inform you about the same danger. In other words, sensory inputs which are temporal and spatial congruent tend to originate from a common source. That is, sensory information that occur at the same place and at the same moment in time tend to integrated together through multisensory integration (MSI) and perceived as one. In pursuance of an effective interaction with a dynamic multisensory environment, the brain imposes order on the multitude of sensory information. Various sensory modalities have to communicate with each other in order to create perceptual representations. Such communications are between the representations of space and time in different sensory modalities. However, space is not a unitary concept.

The spatial reference system is a relational system consisting of located objects, reference objects, and the spatial relations that may obtain between them (Shelton & McNamara, 2001). Based on the underlying coordinate system and the information that is stored in the resultant spatial representation, a distinction can be made between egocentric and allocentric reference systems. Egocentric reference frames define specific locations and orientations with respect to the observer. Allocentric reference frames refer to spatial relations with respect to the environment. Examples of egocentric reference frames include eye-centered, head-centered or body-centered coordinate systems. Egocentric reference frames are sensorimotor representations of space, while allocentric reference frames are map-like representations of space. Integration of different reference frames is crucial for a successful perception and navigation in the environment. Sensory information is

encoded based on an egocentric reference frame in the respective sensory system. Visual, tactile and auditory (as well as vestibular and kinesthetic for navigation in an environment) information is used in efficient manner in the sense that little attention is needed for spatial processing, but any discrepancies between different sensory modalities or attempt to ignore this sensory information is computationally effortful and prone to error. Vision is first encoded in an eye-centered (retinotopic) coordinate system whereas audition in a head-centered (craniotopic) one and somatosensory information are encoded in a body-centered (somatotopic) frame of reference.

Visual space is constructed from a series of retinal images, like snapshots, that are pieced together using information about the direction in which the eyes were pointing in each snapshot to create a 'mind's eye' view of space (Stein, 1989) and are first processed in the primary visual cortex (V1) where each hemisphere processes input from the opposite eye. Auditory space is perceived even more indirectly. The direction of sound source is computed from differences in the amplitude and timing between the two ears, from the coloration provided monaurally by the pinna, and from the direction in which the head and body were pointing at the time. Acoustic inputs are first processed in the primary auditory cortex (A1). Likewise, somatosensory space is a complicated construct, partly built on the somatotopic maps of the skin relayed to the brain, but equally reliant on motor and proprioceptive signals indicating what the limbs where doing when objects were encountered. Tactile information is first processed in the postcentral gyrus, an area also known as the primary somatosensory area (S1). The somatosensory cortex is highly organized, with areas corresponding to specific body parts and the size of these areas depending on the complexity of the sensation processed by that body part.

Yet, all these different spatial coordinate systems are brought together equivalently by the brain into a unified concept of perceived space. The existence of multiple reference frames raises several important questions. How does the brain make sense of incoming sensory information that

are temporally aligned, but initially encoded in a variety of egocentric spatial reference frames? What underlying mechanism enables such spatial perception?

In the following sections, I will begin by reviewing the role of spatial reference frames in the existing literature of MSI. While MSI research includes uncovering the details and limitations of spatial as well as temporal synchrony, for the purpose of this review, I will mostly focus on the spatial aspects of multisensory integration. First, I will describe the different ways that information from different senses can interact on a behavioral level. Then I will describe the neurophysiological perspectives on multisensory integration. After that, I will consider the specific topic of spatial reference frames for MSI. Finally, I will discuss the role of attention in MSI.

3.2. Evidence of multisensory integration in behavior

While the earliest studies on sensory perception investigated sensory modalities in isolation, many researchers have addressed the interactions between modalities in behavioral experiments. Multisensory interactions on one hand may allow for more focus on relevant information, while filtering out background noise. On the other hand, it could also cause distraction when attention is captured involuntarily by task-irrelevant sensory input or biased by another sensory modality.

Intersensory bias is the fast operating process that causes perception to tend to agree to the assumption that one single event has occurred. This can happen even when the spatial locations of the two events are not overlapping. Intersensory bias is a mechanism of the perceptual system in response to a discrepancy of two sensory modalities. Such bias takes place towards the stronger modality, the one that has a higher spatial or temporal resolution in the given task. One example of intersensory bias is the phenomenon of the Ventriloquism effect (Jack & Thurlow, 1973) where the voice of the puppeteer seems to appear from the puppet itself. The effect describes an audio-visual spatial conflict in which vision influences the perceived sound location. The puppeteer is well versed in manipulating the puppet's mouth while moving his lips as little as possible while speaking. Visual

capture occurs, and the audience believes in what they see – a speaking puppet. In terms of reference frames, we can see that while the sound is encoded craniotopically, vision is coded retinotopically. Even though the spatial locations do not match, observers experience a unified concept – that of a speaking puppet. The standard explanation of the Ventriloquist effect is that auditory and visual stimuli occur in close enough temporal and spatial proximity; therefore, the perceptual system assumes that a single event occurred. A translocation of sound towards the visual event seems to be the most ecological solution as the spatial resolution of the visual system is superior to that of the auditory system.

This is similar to what we experience when watching a movie: The sound appears to originate from the speaker's lips, instead of from loudspeakers, the actual audio output device. Some of our perceptual experience appears unimodal. For example, when you are trying to understand what another person is saying, you would think that all the information you are receiving comes from what you hear. However, most of the time, but especially when one listens to someone speaking in a noisy environment, perception is facilitated by what we hear, but also by what we see: the speaker's face and lip movements (Sumby and Pollack, 1954; van Wassenhove, 2013).

Although the visual system may dominate judgments of spatial location in many tasks, it is not to say that vision is unaffected by other sensory modalities through crossmodal perception, which does not just manifest itself in incongruent sensory combinations. For instance, visual detection can be enhanced at the location of sound (Frassinetti et al., 2002) or touch (Macaluso et al., 2000). Sham and colleagues (2000) have demonstrated that sometimes the auditory system can also trick the visual system into an illusion. Participants in their study were asked to hold fixation on a computer monitor and report how frequently a white disk was flashed in the periphery of the screen. Participants were very accurate when performing this purely visual task alone. Notwithstanding, when one of these flashes was accompanied with multiple auditory beeps, participants were fooled into believing they had seen several flashes of the disk, a sound-induced flash illusion. While this

provides ample evidence of what we hear can influence what we see, it needs to be pointed out that the experiment demonstrates the temporal superiority of the auditory system over the visual system. Intersensory biases have been shown to occur between various sensory combinations such as vision and audition (Jack & Thurlow, 1973), audition and touch (Caclin et al., 2002), vision and proprioception (Pavani, Spence & Driver, 2000).

Moving visual stimuli seem to capture the direction of moving sound. This effect is very robust, as it even holds when movements are in fact in opposite directions (Oruc, 2008), but also for different stimuli combinations such as visual-tactile. Caclin et al. (2002) have shown that sound localization judgments can be biased by spatially incongruent, but temporal synchronous tactile stimulation. They found that tactile stimulation to the fingers will capture auditory perception irrespective of spatial attention to the tactile stimuli. Others found that the same holds true, at least for simple auditory and visual stimulus combinations such as beeps and light flashes (Bertelson et al., 2000; Vroomen et al., 2001).

Multisensory integration can also result in surprising new percepts, where no sensory dominance occurs, instead MSI in which the result is not equal to either sense. The archetype of such crossmodal illusions is the McGurk Effect (McGurk & McDonald, 1976), which involves an alteration of meaning. The perceived phoneme is neither, the correct visual nor auditory stimulus by itself, but a new phoneme that derives from a combination of both stimuli information. Integration of sight (the speaker's face and lip movement) and the sound of speech enhance perceptual brain activities (Calvert, Campbell & Brammer, 2000; Sams et al., 1994; van Wassenhove et al., 2007). As discussed earlier, when listening to a speaking person we often do not only rely on what we hear, but also heavily on what we see: their lip movements. Normally, in a real life situation visual and auditory input are congruent, so we integrate them into one percept. It is especially helpful to use visual cues in a noisy environment to decipher ambiguous sounds. When a slight mismatch occurs such as in the McGurk Effect for example the audio of the syllable ["ba"] with the visual labial of the

syllable [“fa”], the resulting synthesis yields a new illusory auditory percept of the syllable [“ga”]. The perceived sounds are often consistent with the visual stimuli in terms of place of articulation (such as labial-lip articulated sounds like [/b/ /m/] or non-labial sounds articulated behind the lips [/d/ /n/]), while also consistent with the auditory stimulus in terms of manner of articulation. It is known that the brain integrates redundant spatial information from vision and audition & provides information about higher-order aspects, here phonemes or as in ventriloquism – about spatial location.

3.3. Neurophysiological perspectives on multisensory integration

It is not surprising that in the past many neuroscience studies have focused on understanding the underlying mechanism of individual senses. Thanks to the advances in neuroimaging techniques over the last years, these sensory areas could be investigated more in depths and detail, and it is now clear that in fact much of the brain does not only consist of specialized unisensory areas, but is also multisensory. As shown in the earlier section on behavioral evidences of MSI it has been amply documented that stimuli from different modalities largely interact. Despite the increasing interest in interactions between the senses, there is a long-standing debate concerning the underlying mechanism of multisensory integration. The concept of primary sensory areas has been well established for decades. These are defined, specific brain regions dedicated to the perception and processing of individual sensory information from one, but not any other, sensory modality. Such unimodal areas, in the occipital cortex for vision, regions in and around the superior temporal gyrus for audition, and regions in the post-central cortex for touch, were thought to communicate with associational areas, which then in turn integrate percepts. This hierarchical model of multisensory integration is depicted in Figure 2a. It shows that incoming sensory stimuli are initially processed in their respective primary sensory areas, (represented as colored triangles in 2a) and from there integrated in higher order areas (represented by the purple hexagon in 2a).

Evidence has been accumulated for early MSI in unitary sensory regions, mostly through intracranial work in primates and cats. For example, Morrel (1972) found that a great portion of neurons in primary visual cortex responded to auditory stimuli. We know now that these individual sensory areas (as well as motor circuits), which are devoted to processing information from a single sense, are highly interconnected, but also respond to input from different sensory modalities (Macaluso et al., 2000, Driver & Noesselt, 2008). Figure 2b represents such early bisensory communication model, in which early interactions between sensory regions exist, that facilitate an integration of stimuli from different sensory modalities. Imaging data has provided further evidence for early integration effects of audio-tactile as well as audio-visual stimuli in the auditory cortex (Kayser et al., 2005; Kayser et al., 2008; Schurmann et al., 2006) as well as visual-tactile integration in the somatosensory cortex (Kida et al., 2007).

Finally, the alternative account of a mixed model as illustrated in Figure 2c arises from a combination of the first two models. That is, primary sensory regions can be modulated by other sensory stimuli, and are then integrated into precise representations in anatomically higher-order areas.

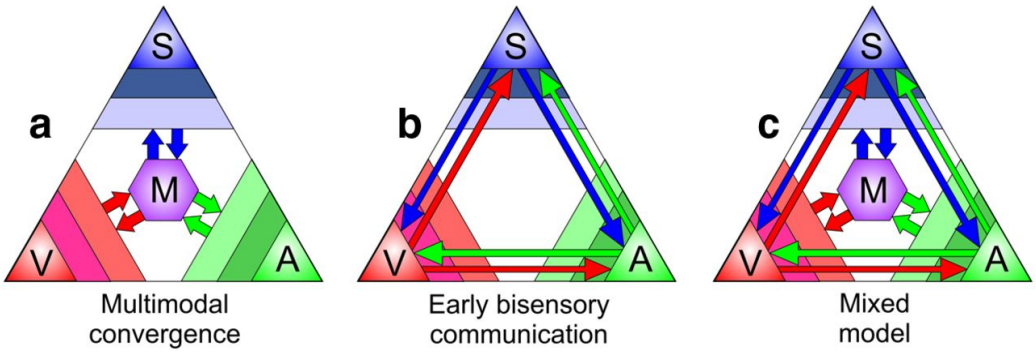

Figure 2. Models of multisensory processing. The three points of the largest triangle represent the somatosensory (S), visual (V), and auditory (A) streams. Color slices in the corner indicate primary, secondary, and tertiary unimodal areas. Hierarchical multisensory processing is represented by the purple hexagon (M). Arrows indicate flow of sensory information. **a:** Traditional, hierarchical views of multisensory processing posit that streams are integrated after initial unisensory processing. **b:** Recent research has also supported the model of early direct interactions between primary sensory cortices. **c:** The mixed model of multisensory processing combines these two views to describe how early unisensory processing can be modulated by other sensory inputs, and later sensory streams can be integrated into spatially precise higher-order multisensory representations. Figure reproduced from Quinn et al., 2014.

The landmark study by Stein in cats (Stein and Meredith, 1993), and later in macaque monkeys (Wallace and Stein, 2001), has shown the involvement of the superior colliculus (SC). This midbrain structure controls orientation and coordination of movements. Furthermore, the superior colliculus receives visual, auditory and somatosensory input, due to the presence of multisensory neurons. Many of these inputs then converge topographically ordered onto a single multisensory neuron (Stein and Meredith, 1993). Ergo, one suggestion of how different spatial reference frames are interpreted by the brain is through the existence of multisensory neurons. These neurons are excited by inputs from more than one sensory modality. Consequently, when multisensory stimuli are present, these neurons can integrate their responses in a predictable manner. Cell recordings in the SC enabled an operational definition of MSI as the presence of a number of neuronal impulses significantly higher (superadditive) when evoked by a crossmodal combination of stimuli, than when evoked by the most effective of these stimuli alone.

These findings helped understanding and establishing the three core principles of MSI. One: MSI is more likely to occur when multiple unisensory stimuli occur at the same location in space. Indeed, multisensory neurons in SC possess diverse receptive fields, one for each modality. These receptive fields overlap in space, and, if two or more stimuli are in the same spatial register, brain activity is enhanced. Consequently, if stimuli are not aligned in space, no enhancement will take place.

Two: As discussed in the beginning of this review, for MSI to take place effectively, stimuli need to be also in close temporal proximity (Bushara, Grafman& Hallett, 2001; Wallace et al., 2004). This is generally defined in time windows, ranging in the order of a few up to several hundreds of milliseconds depending on the complexity of the to-be-integrated stimuli (Spence and Squire, 2003). The existence of such a flexible long time windows makes sense taking into consideration that

different modalities travel at different speed (for example, during a thunderstorm, we see the lightening before the thunder until it is within reach).

Three: Another important principle of MSI is the rule of inverse effectiveness. MSI is more likely to occur when the individual stimuli are rather weak. Very salient inputs are easily detectable, so a combination of them would have a moderate detection and/or localization effect. In contrast, a combination of weaker cues evokes a substantially enhanced response, when compared to them on their own. These findings in animal cell recordings established a foundation for the understanding of MSI in humans as well.

It has been found that between 10-50 % of the neuron in unisensory regions also receive inputs from other sensory modalities (Lemus et al., 2010; Meredith et al., 2012). The discovery of these multisensory neurons allowed for a better understanding of how the brain is organized and helps explain how individuals who suffer a loss of one sense early in life often develop greater acuity in their remaining senses. For example, in an fMRI study, Karns et al. (2012) demonstrated that congenitally deafness can facilitate neuroplastic changes in the human brain. When the auditory cortex receives no auditory stimuli, vision and somatosensation make use of Heschl's Gyrus to process these alternative sensory stimuli. Compared to a control group, the Heschl's gyrus of congenitally deaf individuals showed an increased response to somatosensory stimuli as well as, albeit to a lesser degree, visual stimuli due to the existence of multisensory neurons (Karns et al., 2012). Neuroimaging as well as single-cell recording and behavioral data have brought to light that the relative locations, intensity and the timings of stimuli determine multimodal perception in the brain (Andersen et al., 1997; Felch et al., 2016; Wallace et al., 2004). Furthermore, stimuli that are temporally and spatially aligned can elicit a multimodal response that exceeds unimodal responses (Macaluso et al., 2000; Stein et al., 2004).

In a study using positron-emission tomography (PET), Macaluso et al. (2000) found that mechanisms of sustained spatial attention in vision and touch operate at modality-specific, but also

on multimodal levels. Participants had to attend either to a series of light flashes or a series of finger vibrations presented either to the left or to the right periphery. The researchers sought out to identify brain areas showing differential activations to attended hemifield, but also to distinguish areas showing this spatial effect within only one modality versus multimodal effects, a spatial effect independent from stimulated modality. In the first experiment, subjects were instructed to attend to a modality and a side (vision-attend right, vision-attend left, touch-attend right, touch-attend left). Activations for the visual task fell within cortical areas involved in low-level visual judgments (BA 19), such as the anterior medial occipital gyrus, the superior occipital gyrus and the superior parietal lobe, whereas tactile stimuli caused activation in cortical areas previously associated with somatosensory processing – the inferior postcentral gyrus and superior postcentral gyrus. Modality-specific attentional effects (right vs. left) were found in the left superior occipital gyrus for visual tasks, and in the left superior postcentral gyrus for somatosensory stimuli. Their second experiment only involved tactile conditions, but manipulated the presence of vision (eyes open versus eyes closed), and replicated the findings of experiment one. Interestingly, they also found activation in the left intraparietal sulcus only in the eyes open condition (when participants could see their stimulated hand), which confirms the multimodal nature of the cortical area.

3.4. The importance of spatial reference frames

On one hand, we humans have a strong naïve impression of unified space. On the other hand, laboratory studies show that we are easily fooled and often misperceive or have incoherent perceptions. This might reflect the fact that our brain codes the location of stimuli in multiple different reference frames simultaneously, and it is not clear how, or even whether these different reference frames are combined at any moment in time. Various different solutions have been proposed such as the supra-modal reference frame. Here, sensory reference frames are translated into something new that consists of a combination of the individual coordinate systems. For instance,

orienting towards a multisensory target involves, at the input stage, a sensory code in a retinotopic map for the visual cues, in a head-centered map for the auditory cues, and in a somatotopic map for the tactile cues. A constant transforming of these diverse maps into a common frame would be computationally heavy as environment is not static, especially during locomotion. An allocentric reference frame must then be created anew constantly, as it is not one of the sensory reference frames. This solution seems to be too costly and therefore very unlikely as spatial processing of known and new environments alike seem to be automated processes that require little attention. Some studies propose that spatial information is remapped between different frames of reference. Although there are some differences in implementations, various theoretical accounts posit a transformation process from one reference frame to another implying a space-to/from-body remapping process (Heed et al., 2015, Badde et al., 2014, Noel and Wallace, 2016; Zampini et al., 2003). Transformations that translate for example auditory into retinotopic coordinates still remain as the same distinct reference frames, but are re-coded. Exactly how and where such remapping takes place in the human brain remains to be uncovered. A third proposed solution is an alignment of distinct unisensory reference frames (Stein, 1992). How would these individual coordinate systems align in order to accurately locate and act upon a multisensory stimulus in space?

As discussed earlier Stein and Meredith (1993) established the involvement of multisensory neurons in MSI. When visual, auditory and somatosensory inputs are received, many of these converge topographically ordered onto a single multisensory neuron (Stein and Meredith, 1993). One suggestion of how different spatial reference frames are interpreted in the brain is through these multisensory neurons. These neurons are excited by inputs from more than one sensory modality. Therefore, when multisensory stimuli are present, these neurons can integrate their responses in a predictable manner.

In contrast, Spence and Driver (2004) argue for a transformation of the initial natural reference frames into a common predominantly eye-centered frame of reference. Their argument is

based on the superior spatial resolution of the visual system compared to the other sensory systems. Thus, the superior colliculus represents stimulus position predominantly with respect to the current gaze position rather than to its position in sensory space. This is plausible as this brain area is closely tied to eye movements and orienting responses, and is made up of mostly oculomotor neurons. Cohen and Andersen (2002) found further support of this theory through a series of reaching experiments in monkeys. They showed that parietal neurons in the posterior parietal cortex transform sensory eye-, head-, body-, or limb position signals into a common eye-centered frame of reference. Such common reference frame in turn can facilitate movement coordination communication and might thus be an efficient way to represent locations of different sensory targets. Based on behavioral findings in transcranial magnetic stimulations (TMS) to the posterior parietal cortex (PPC), Pasalar et al. (2010) confirmed this brain region's involvement in visual-tactile multisensory integration.

Several other studies have since confirmed that establishing an egocentric reference frame through integration of multiple sensory inputs is one of the many tasks of the posterior parietal cortex (Calton and Taube, 2009; Buchholz et al., 2011; McGuire and Sabes, 2011; van Gilder et al., 2016; Kaulmann et al., 2017). However, many details about these spatial reference frames remain an issue of much debate and two theories have crystalized: one being that frames of reference exhibit temporal development, suggesting a dynamic evolvement of coordinate transformations in a large neuronal population. That is, neurons encode information in different reference frames at different times throughout a task. Per contra, frames of reference show no temporal development and therefore different neuronal populations encode coordinate transformations, which are necessary for response movements etc., simultaneously across the parietofrontal circuit (Buneo et al., 2008; Hadjidimitrakis et al., 2017). With their publication, Hadjidimitrakis et al. (2017) are in support of the later theory as they found that most V6A neurons in monkeys consistently used body-centered frames of reference through the chronological stages of a delayed reaching task. Although at earlier stages of the task, some neurons utilized mixed body-, hand-centered reference frames, which could suggest an early

involvement of these neurons in eye-hand coordination. Whereas the preeminent pure body-centered frame of reference might be indicative of online movement control (Hadjidimitrakis et al. 2017).

A majority of studies investigating spatial reference frames has been of unisensory nature: stimuli consisted of one single modality, and often included executing a movement. To date, much is still unknown about remapping of spatial reference frames from different sensory modalities. If stimuli are truly remapped into an eye-centered frame of reference, what happens when vision is taken out of the game? A remapping to a retinotopic reference frame seems to makes little sense in the blind or when integrating sensory inputs that does not involve vision. What then happens when vision does not facilitate MSI, can it hinder it? Interesting findings come from MEG results in a blind subject (Ioannides et al., 2013). In response to tactile stimulation to the median nerves, the researchers found significantly different time-frequency differences from the alpha band up to higher frequencies in the visual cortex of the blind participant compared to two control participants. Based on time-delayed mutual information, Ioannides and colleagues (2013) further introduced first MEG evidence that in the blind, somatosensory information is funneled from primary somatosensory cortex through posterior parietal cortex to visual brain areas.

Various bimodal experiments have brought to light that the spatial link in MSI between the auditory and somatosensory modalities is much weaker than the one between other modality parings. Different research groups found spatial modulatory effects on temporal order judgment tasks (TOJ) regarding which modality was presented first between auditory and visual stimuli (Spence et al., 2003; Zampini et al., 2003), and between visual and tactile stimuli (Spence et al., 2003). That is participants' performance was worse when the two stimuli (auditory and visual; visual and tactile) were presented in the same location, compared to opposite locations. Interestingly, Zampini and colleagues (Zampini et al., 2003) found that spatial TOJs ("Which side was presented first?") were unaffected by the stimuli modalities. Through a series of experiments, they found that performance was better for modality TOJ than for spatial TOJs. These results revealed that the relative spatial

reference frames of the presented stimuli as well as the particular stimulus dimensions bear a critical influence on MSI. Spence and colleagues used an interesting variation of the TOJ task by adding a crossed-hands condition and found that this caused a decline in performance, suggesting a failure to remap visuo-tactile space accurately (Spence et al., 2003; Shore, Spry & Spence, 2002). In contrast, such spatial modulatory effects, as seen in the visuo-tactile and audio-visual domain, seem to be absent in audio-tactile TOJs (Zampini et al., 2005). Murray et al. (2005) reported similar results in an EEG study. They found no differences in reaction time facilitation or distinguishable neural correlates when audio-tactile stimuli were presented either right/left aligned or misaligned. They furthermore localized audio-tactile interactions to auditory association areas contralateral to the side of somatosensory stimulation at around 50 ms post-stimulus onset. They concluded that audio-tactile integration occurs early in sensory processing in a traditionally considered auditory (and thus unisensory) area and across a wide spatial separation. The authors themselves advocate that this holds true for early audio-tactile MSI across space, and that perceptual-cognitive phenomena such as capture and ventriloquism manifest at later stages of sensory processing (Murray et al., 2005).

However, recently Noel and Wallace (2016) made use of TOJ task in order to further investigate the influence of conflicted somatotopic and external reference frames by utilizing different body postures as well as conditions of sensory deprivation. Participants were asked to judge the order of two tactile stimuli to their ankles in conditions where their legs were either uncrossed or crossed, which results in conflicting somatotopic and external reference frames. As expected, performance was worse in the crossed leg condition. A possible explanation is that the primary somatosensory cortex maps skin location independently of the posture of a give body part (Penfield and Rasmussen, 1950; Hlushchuk and Hari, 2006), so in order to make sense of a tactile stimulus the brain must realign tactile coordinates in order to locate the origin of that stimulus (Heed et al., 2015). Such remapping takes place by combining somatosensory input with proprioceptive and visual cues about body postures (Soto-Faraco, Ronald & Spence, 2004). An early somatosensory activity arises

from the initial feed-forward sweep of neural activity to the primary somatosensory cortex. The later conscious experience is brought to life by the activity of the somatosensory network involving recurrent connections from association areas (Azañón & Soto-Faraco, 2008). Additionally, in Noel and Wallace‘s study participants were deprived of vision by blindfolding and/or placed in an anechoic chamber to remove all auditory input. Visual deprivation alone showed no significant effect on the task performance differences between uncrossed and crossed leg postures. However, auditory and even more so, audio-visual deprivation intensified the tactile temporal acuity differences. These results reveal that the most detrimental modulation of tactile localization is the combined removal of audio-visual information. Hence, both audio and audio-visual spatial information are remapped onto and integrated with somatosensory anatomical locations in the context of spatial localization. However, what seems to be surprising is the directionality of the effect. Given the modality-specific spatial reference frames, the removal of another sensory modality should make spatial localization easier and not hinder it. Furthermore, why did visual deprivation, the removal of the visual spatial reference frame show so little effect here? Can we explain it based on Spence’s and Driver’s translocation of reference frames into the retinotopic frame of reference?

One idea is that visual and auditory stimuli that appear near the body or on a particular body part (“near space”) are mapped in a body-centered fashion (Graziano et al., 1997; Graziano et al., 1999; di Pellegrino & Làdavas, 2015), and thus provide redundant information rather than conflicting information. A removal of this integrated remapped information then causes detrimental effects. These results indicate that other sensory modalities and the combinations thereof, have an influence to the mechanisms governing alignment between somatotopic and external reference frames. They strongly suggest an alignment of multiple reference frames takes place in order to decipher exteroceptive spatial information. Taken together, these TOJ effects indicate that redundant spatial cues can facilitate MSI. It becomes clear that for MSI to be effective spatial as well as temporal factors play a crucial role, even suggest an intimate link between them. The above studies further

suggest that the magnitude of MSI depends on the particular combination of sensory modalities, and propose that audio-tactile interactions may be less spatial than any sensory modality combination in MSI involving vision.

Heed et al. (2015) suggest that integration of multiple concurrently active spatial representations best describes the process of spatial localization. They propose coordinated oscillatory activity as a suitable underlying mechanism that allows for large-scale parallel representations of multiple spatial formats and the formation of an integrated location estimate. Although studies have shown that a translocation and alignment of reference frames takes place, they are limited in terms of investigating the interplay of just two sensory modalities.

3.5. Is multisensory integration automatic or strategic?

An important topic to consider is whether MSI is an automatic process and therefore always occurs. An automatic process is insensitive to the load of the current task demands and is consequently not influenced by other competing sensory events. In addition, an automatic process has to satisfy the intentionality criterion. This means that MSI should always occur irrespective of the observer's voluntary goal control (top-down control).

The McGurk effect along with many other multisensory illusions suggests that attention plays very little of a role in multisensory integration. Instead, these effects seem to reflect largely automatic sensory interactions. Audio-visual association of facial gestures and vocal sounds has been demonstrated in nonhuman primates and prelingual children. This argues for a general basis for this capacity, which does not require much attention. This is interesting as the functions of spatial cognitive processing are located in cortical regions that are strikingly similar in different species indicative of an evolutionary old system central to the survival of the a species. These cortical areas of the spatial network are heavily interconnected and built the basis for the computation and maintenance of spatial information based on distinct reference frames (Gramann, 2013).

However, over the recent years various experimental variations of the McGurk effect have shown that the effect breaks down when attentional resources are allocated elsewhere (Alsius et al., 2005; Soto-Faraco & Alsius, 2007; Tiippana et al., 2004). For example, Alsius and colleagues (Alsius et al., 2005) have demonstrated that when observers' attention resources are depleted through additional, simultaneous but unrelated visual or auditory tasks the McGurk effect breaks down. This challenges the idea that crossmodal speech integration is automatic. Which in turn, implicates that attention is needed for such complex multisensory phenomena by binding features across modalities. Interestingly, Japanese (and Chinese) listeners are less susceptible to the effect (Sekiyama &Tohkura, 1993; Sekiyama, 1997). This hints at the idea of cultural influence. Japanese listeners seem to rely less on facial information during speech perception. This is in line with the notion that spatial reference frames rely on the environment during perceptual development. The neural basis for cortical regions is genetically determined. This results in soft-wired neural structures that adapt to cognitive activity (Gramann, 2013). It is thus possible that during speech perception Japanese listeners rely dominantly on auditory information. This may be rooted in Japanese culture, as it is considered rude to look at a person's face, especially a person of higher social status, during communication.

Intriguingly, Oruc et al. (2008) demonstrated with the crossmodal dynamic capture paradigm that attention does have an effect on how different motion signals are combined, but only when the susceptibility for capture between the two modalities is comparable. They investigated bimodal apparent motion streams consisting of a pairing of visual, tactile or auditory stimuli, where one modality represented the target, while a second modality acted as a distractor. Participants had to discriminate the direction of a target stream (visual, auditory or tactile) while trying to ignore the direction of a distractor stream presented in a different modality. When target-distractor pairings were blocked or cued at the beginning of each trial, then visual motion captured synchronous auditory or tactile motion. Nevertheless, neither auditory nor tactile stimuli could capture visual

motion, while synchronous auditory and tactile stimuli could capture each other. However, when no attention was directed prior to stimulation (participants did not know which was the target and which was the distractor modality until after stimuli presentation), participants misperceived motion direction in auditory-tactile stimuli combinations, while performance for pairings that involved vision remained the same. The visual system appears to be unaffected by incongruent crossmodal distractor stimuli in motion judgment tasks. This is likely due to the superior spatial localization capabilities of the visual stream.

CHAPTER 2: ALPHA FREQUENCY AS A PREDICTOR OF PERCEPTUAL OUTCOME IN TEMPORAL INTEGRATION

Introduction

As detailed in the previous chapter, alpha frequency oscillations play an important role in the temporal dynamics of stimulus perception and processing. A recently emerging corpus of studies has investigated the relationship between individual alpha peak frequency and the temporal resolution in sensory perception (Samaha and Postle, 2015; Baumgarten et al., 2017). Based on EEG finding of a two-flash fusion threshold paradigm, Samaha and Postle (2015) offered supporting evidence that individuals with higher alpha frequency have finer temporal resolution in vision (Figure 3), which suggests that visual alpha represents a mechanism for temporal discrimination of successive visual stimuli (Samaha and Postle, 2015).

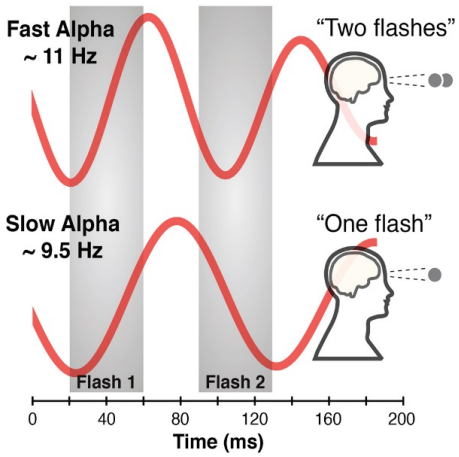

Figure 3. Individual alpha frequency. Individuals with higher alpha frequency have finer temporal resolution and thus can successfully segregate two successive presented flashes. Figure reproduced from Samaha and Postle, 2015.

In Samaha and Postle's experimental paradigm, either two brief visual stimuli (40 ms each) were presented in the same spatial location and are separated by a variable ISI (10-50 ms in steps of 10 ms), or one visual stimulus was presented for identical overall stimulus duration (90-130 ms). Subjects had to respond if they perceived one or two visual flashes. On group-level, performance improved with increasing ISI in the two flash conditions. In contrast, group-level performance in the one flash condition showed little target duration effects. The authors then showed that the individual length of ISI necessary for the subject to correctly segregate the two flashes correlated with each subject's individual occipital alpha frequency. To this end, they derived the individual alpha frequency for each subject during an eyes closed EEG recording as well as during a pre-stimulus period and observed that correlational pattern in both cases. They found that faster alpha frequencies predicted more accurate flash discrimination, while theta and beta frequencies did not. Furthermore, the authors found that within-subjects pre-stimulus higher instantaneous alpha frequency preceded correctly discriminated trials. However, individual differences in alpha power did not predict fusion thresholds. Recently, Baumgarten et al. (2015; 2017) investigated if these findings were also applicable in the tactile domain by means of MEG recordings. They found through differences in phase, that the beta band (13-30 Hz) took on the role of temporal integration windows in the somatosensory domain, but failed to find any correlation between individual alpha or beta peak frequencies and perceptual outcome.

As described in Chapter 1 of this thesis, alpha power is reflective of the state of cortical excitability through functional inhibition (Jensen and Mazaheri, 2010; Foxe and Snyder, 2011). In general, shifts of attention in space, towards a sensory modality or a specific stimulus feature are reflected in pre-stimulus alpha band power. Thus, it seems rational that the state of pre-stimulus alpha could determine the perceptual outcome of a near-threshold stimulus in a paired-stimulus paradigm. Near-threshold studies in the tactile (Schubert et al., 2009; Weisz et al., 2014), auditory (Leske et al., 2015) and visual domain (Ergenoglu et al., 2004; Romei et al., 2008a; van Dijk et al.,

2008; Lange et al., 2013) revealed that stimulus perception was influenced by alpha power (and phase, as described earlier) in the time period preceding the onset of the stimuli.

For example, with their publication, Schubert et al. (2009) set out to understand if conscious perception is determined by ongoing brain states. While recording EEG data, they used a somatosensory backward masking paradigm in which participants were always to respond to weak near-threshold stimuli on one hand (target), irrespective of if it was presented alone or with an ensuing above threshold-stimuli on the other hand (mask). In absence of a target, however they were asked to respond to the mask stimuli. Target perception was accompanied by pre-stimulus beta frequency desynchronization over frontal cortex (contralateral to mask hand) and a subsequent pre-stimulus attenuation of both 10 Hz and beta in the respective somatosensory cortices. Interestingly, the researchers also found a strong correlation between participants' target perception and their individual frequency amplitude in those cortical areas. They interpreted the strong contralateral beta desynchronization to be reflective of top-down control in the prevention of backwards masking. Like in other above-mentioned studies, perceptual outcome differences (detection versus misses) of near-threshold stimuli were characterized by a relative decrease in alpha (and beta) power contralateral in the respective primary sensory region prior to successful stimulus detection. Such a decrease can be interpreted within the functional inhibition framework as a release of functional inhibition of task relevant areas, making stimulus processing more efficient and conscious perception more likely. In turn, the comparison of evoked responses of different stimulus percepts e.g. near-threshold stimulus perceived or missed enables researchers to further explore neuronal correlates of conscious perception in sensory cortices.

Motivation for the present study

Given the idea that alpha acts as the gatekeeper of temporal integration windows, the paired-stimulus paradigm with a near-threshold stimulus lends itself as a suitable method in further

understanding the precise temporal dynamics. If the two stimuli appear in rapid succession, they are likely perceived as one event. With increasing temporal distance between the two stimuli, such integrated perception will atrophy and the flash events will be perceived segregated in their true quantity. Given suitable timing, identical paired-stimulus presentations may result in different outcomes (integration or segregation) on a trial-by-trial basis. To this end, we used an interesting variation of the paired-stimulus paradigm that enabled us to scientifically investigate oscillatory parameter in the pre-stimulus time period, as well as different perceptual outcomes.

Materials and Methods

Participants

20 subjects participated in the experiment (9 female, mean age= 25.6 years, SD = 2.39 years, all right-handed). All participants provided informed consent as approved by the institutional ethics committee took part in exchange for monetary reimbursement and had normal or corrected-to-normal vision. The study had prior approval by the Ethical Committee of the University of Trento and was conducted in accordance with the Declaration of Helsinki as revised in October 2008. Two subjects were excluded from the analysis due to magnetic inference suspected to the resultant of unreported dental work. One subject aborted the experiment prematurely.

Task and design

On each trial, two localized visual flashes were presented on a medium (50 %) gray background (see Figure 2A). Flashes were 2D Gaussian luminance distributions with an approximate total diameter of 1 degree visual angle, although the perceived size of the Gaussian blob may have been smaller depending on subject and adjusted contrast. Flashes were shown at two contrast levels: full contrast and threshold contrast (Figure 4.B). The duration of each flash was set to 8.3 ms (one screen refresh, 120 Hz). In full contrast flashes, the peak of the Gaussian blob was white (maximum luminance),

giving maximum luminance contrast against the background. Threshold contrast flashes were adjusted by a Quest procedure (Watson & Pelli, 1983) until subjects were just above chance performance (57 % correct) at detecting a single threshold flash. To reduce the number of chance hits included in the analysis, flashes (or pairs of flashes) were shown in one of four quadrants, symmetrically arranged at a diagonal distance of 6 degrees visual angle around the point of fixation, and randomly chosen for each trial (Figure 4.A). Subjects were then asked to identify the quadrant where they had seen the flash, rather than whether they had seen it.

Each trial was chosen from one of 6 experimental conditions in a randomized fashion. Four dual-flash conditions with different ISIs (33, 67, 100 and 400 ms) were intended to probe the temporal integration behavior of the subjects. Additionally, a single threshold pulse and two full contrast pulses with an ISI of 33 ms were included as catch trial conditions.

After each trial, subjects were prompted to make two responses. The first asked to indicate the screen quadrant in which the flashes were perceived. The second asked to indicate the number of flashes perceived (one or two). Responses were made by means of an MEG-compatible button box (ResponsePixx), with matching color coding between buttons and displayed questions (Figure 4.A). As illustrated in Figure 4.A., each trial started with a central fixation cross (black, 1 ° visual angle on a uniformly grey background for a variable pre-stimulus interval (500-1000 ms, randomly distributed) followed by the experimental trial presentation in which the flashes were presented. After a variable post-stimulus interval (500-1000 ms, randomly distributed), participants were asked to make their responses.

Participants were instructed to maintain fixation on the central fixation cross throughout the experimental trials in order to minimize eye movement artifacts.

After completion of the main experiment, which consisted of 5 blocks of 232 trials, subjects were directed to look at a fixation cross at the center of the screen, without further tasks. During this, 5 min of MEG data was recorded for the purpose of resting state analysis.

The stimuli were generated on an HP Intel Quad core computer using Matlab 8.0 (MathWorks) and Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997) and presented to subjects via back-projection by a DLP projector (Panasonic PT-D7700E) on a translucent screen at a refresh rate of 120 Hz in a dimly lit magnetically shielded room (AK3B, Vakuumschmelze, Hanau, Germany).

The precise timing of the visual stimulation was calibrated via a photo diode placed at the upper left corner of the projection screen and controlled with a data and video processing peripheral (DATAPixx, VPixx Technologies Inc., Saint-Bruno, QC, Canada).

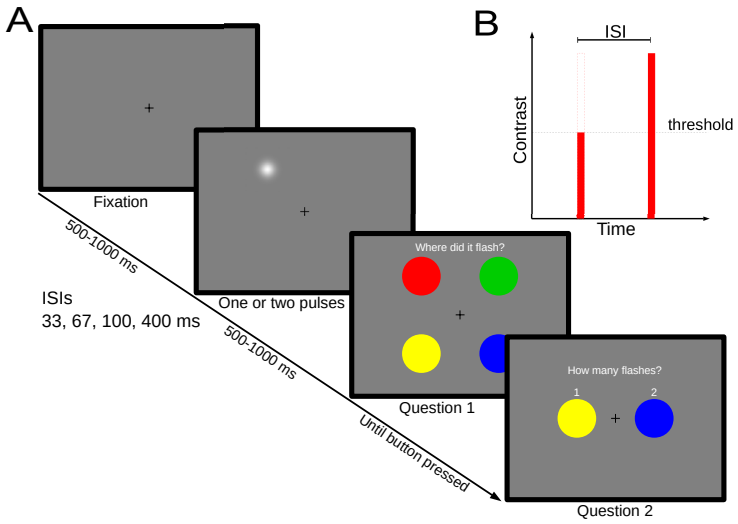

Figure 4. Illustration of dual flash experimental trial. **A:** After a variable inter-trial interval during which participants fixated on a central fixation cross, two flashes of light were presented in rapid succession at different ISIs (33, 66, 100 and 400 ms). Participants were to first indicate in which quadrant the stimuli appeared and then the number of flashes they perceived by pressing color-coded buttons. **B:** The first stimulus was presented at threshold intensity, while the second stimulus was presented at above-threshold intensity.

Data acquisition

MEG data were recorded at a sampling rate of 10 kHz using a 306-channel (204 first order planar gradiometers, 102 magnetometers) VectorView MEG system (Elekta-Neuromag Ltd., Helsinki, Finland) in a magnetically shielded room (AK3B, Vakuumschmelze, Hanau, Germany). Hardware filters were adjusted to band-pass the MEG signal in the frequency range of 0.01–1,000 Hz. Prior to the experimental runs, a subject-specific head-frame coordinate reference was defined, in order to

localize the head position of the subject within the MEG helmet. The cardinal points of the head (nasion, left and right pre-auricular points), the location of five head-position indicator (HPI) coils, and a minimum of additional 200 head-shape samples were digitized for motion tracking (3Space Fastrack; Polhemus, Colchester, VT) at the start of each session. The subject's head position relative to the HPI coils and the MEG sensors was estimated before each experimental run to ensure that no large movements occurred during the data-acquisition procedure.

Data preprocessing

Environmental noises were removed and the data was co-registered in order to remove small head movements across the separate measurement runs through Signal Space Separation with spatio-temporal extension (tSSS; Taulu & Hari, 2009) implemented through the MaxFilter software version 2.2.15 (Elektra-Neuromag Ltd., Helsinki, Finland). Prior to that, data was visually inspected and noisy channels were excluded from the tSSS maxfiltering.

Data were then analyzed using the Fieldtrip toolbox (Oostenveld et al., 2010), the CoSMoMVPA toolbox (Oosterhof et al., 2016) and custom-built Matlab functions in combination with Matlab 8.1 (MathWorks, Natick, MA). Data was downsampled to 250 Hz and 40 Hz lowpass-filtered. Epochs of 4 s (2 s pre and 2 s post) were centered on the stimuli onset. Zero seconds therefore represent the onset of the first stimulus. Trials were visually inspected for additional possible artifacts and contaminated trials were excluded from further processing. Across subjects an average of 1016 trials (SD= 241) of the 1160 trials were retained. The continuous resting state recordings were cut into non-overlapping 10000 ms epochs and further processed in identical manner as experimental trial data.

Data Analysis

Time-frequency analysis in sensor space

Experimental data was analyzed in the frequency domain between 2-30 Hz (in steps of 1 Hz). An adaptive sliding window with a length (Δt) of 5 cycles of the respective frequency was applied and shifted in steps of 10 ms between –1.0 to 0.5 s. A Hanning taper was applied, yielding a spectral smoothing of 1/Δt. The tapered epochs were subsequently Fourier-transformed in a 2-30 Hz range and the power estimates averaged over trials. A threshold-free cluster enhancement test (tfce; Oosterhof et al., 2016) that efficiently handles the issue of multiple comparisons was applied for statistical testing of significance at occipital-parietal sensors corresponding to primary visual cortex from -500 ms to stimulus onset. In particular, for each ISI spectral estimates between 8 and 14 Hz were averaged and hits and misses were contrasted via tfce-based one-sample t-statistics against zero with Monte-Carlo randomization.

Individual alpha frequencies

Resting state epochs were linearly detrended, multiplied by a Hamming window, zero padded and fast Fourier transformed. The resulting power values were extracted and IAF was identified as the local maximum within the frequency range of 7 to 13 Hz. Each subject showed a clear peak within this range in all sensors (see Figure 5.B for a sample subject). Based on different performance improvements between increasing temporal gaps between flashes across subjects, we asked whether their IAF was correlated with such performance improvement. For each subject, mean performance differences between subsequent ISI were calculated. Correlations between IAFs and mean performance differences were assessed by means of Pearson correlation. Topographic distributions of these correlations were generated by applying a Spearman correlation, as it is less sensitive to potential outliers.

Instantaneous frequency

Experimental data were first band-pass filtered with a zero-phase, plateau-shaped FIR filter in the alpha frequency band between 7 and 13 Hz. Then we computed the instantaneous phase angle over time with a Hilbert transformation. As detailed in the publication by Cohen (2014), the instantaneous frequency is defined as the time rate of change of the instantaneous phase angle. Thus, the temporal derivate of the instantaneous Hilbert phase corresponds to the instantaneous frequency in Hertz (scaled by sampling rate and 2π). As the resulting phase angle time series is prone to noise that could cause sharp artifacts, we applied a median filter (10 equally spaced window sizes between 10 and 400 ms) for ten times. Across those resulting median-filter windows, we calculated the median instantaneous frequency estimates. As task performance was comparable at ISIs of 67 ms and 100 ms, we pooled this data together and focused our analysis on subjects that reached at least 20 % of accuracy (N= 14) in order to obtain adequate statistical power. For each perceptual outcome and for each subject an equal number of trials (in which participants had first correctly identified the correct stimulus location) were randomly selected to prevent any bias across conditions.

We then statistically compared the difference in instantaneous alpha frequency between consciously perceiving one or two flashes. Following the methodology of Samaha and Postle (2015), we selected the sensors with the highest pre-stimulus power (right occipital: MEG2511 and MEG2541) and then calculated the difference between perceptual outcomes in the pre-stimulus time period from -500 ms to 0 ms (onset of the 1st flash stimulus) with a dependent samples t-test with correction for multiple comparison by means of an nonparametric cluster-based permutation procedure.

Results

Behavior

Accuracy was measured as the proportion of correct responses of total trials for each ISI. A response was counted as correct when participants correctly identified the quadrant in which the stimuli appeared and furthermore correctly indicated the number of stimulus flashes (1 or 2).

Across participants (N = 16) mean accuracy was 23.53 ± 4.62 % at ISI =33 ms, 36.84 ± 6.37% at ISI = 67 ms, 39.71 ± 5.98 % at ISI = 100 ms and 50.93 ± 3.04 % at ISI = 400 ms. A one-way analysis of variance between the performance at different ISI revealed significant statistical differences, $F(3, 56) = 4.75$, $p= 0.0051$. Further multiple comparisons revealed that performance at the longest ISI of 400 ms was significantly better than performance at the shortest ISI of 33 ms (mean improvement of 27.4 %, $p= 0.0023$). Figure 5.A clearly shows this trend of behavioral improvement with increasing ISI.

Taking a closer look at individual subject's performance, displayed as individual grey lines in Figure 5.A, revealed that subjects differed in their ability to correctly discriminate the two flashes. While some subjects already showed performance improvements when the ISI increased from the lowest of 33 ms to 67 ms, others reached their best performance only at the longest ISI of 400 ms.

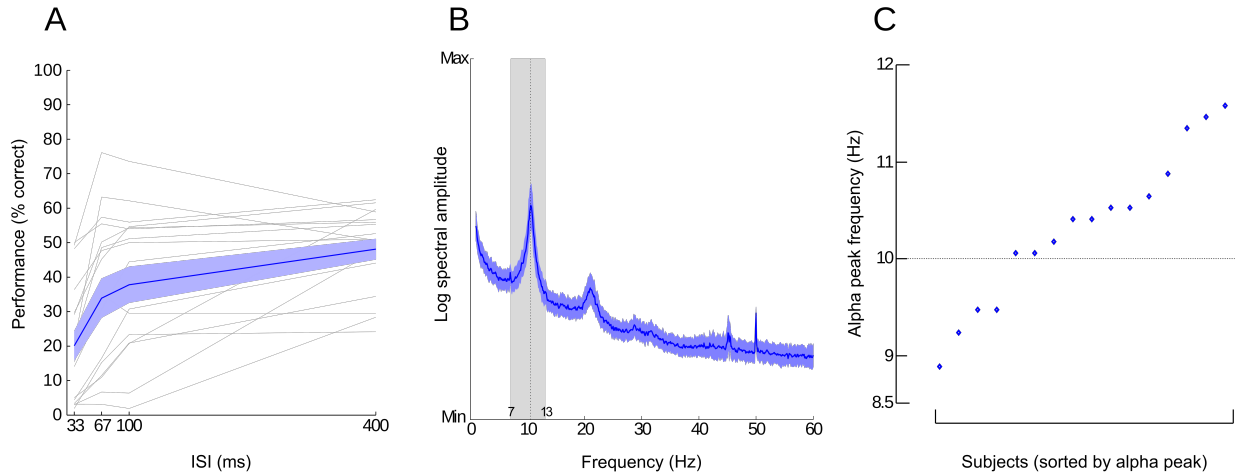

Figure 5. Behavior and Resting State analysis. **A.** Behavioral performance: Blue line as mean and standard error of performance as a function of ISI, averaged across subjects (N= 16, subjects with resting state recordings). Grey lines depict individual subjects' performance as a function of ISIs. **B.** Resting state spectrum of a sample subject (magnetometers): dashed line depicts peak alpha frequency within alpha range (grey shaded 7-13 Hz). **C**. Alpha peak distribution: Results of IAF analysis. Subjects' individual alpha frequency showed variability with the alpha frequency band. Horizontal insert gives group mean.

Time-frequency analysis

Following up on the behavioral findings that showed how behavioral improvement varied between subjects, we had hypothesized to find a correlation between individual performance improvement and individual alpha frequency.

As shown in Figure 5.B. for a sample subject, IAP was clearly identifiable within the alpha frequency range of the resting state spectrum. Individual alpha peak frequencies were clearly identifiable in all subjects and ranged from 8.9 to 11.55 Hz (mean= 10.35 Hz; Figure 5.C).

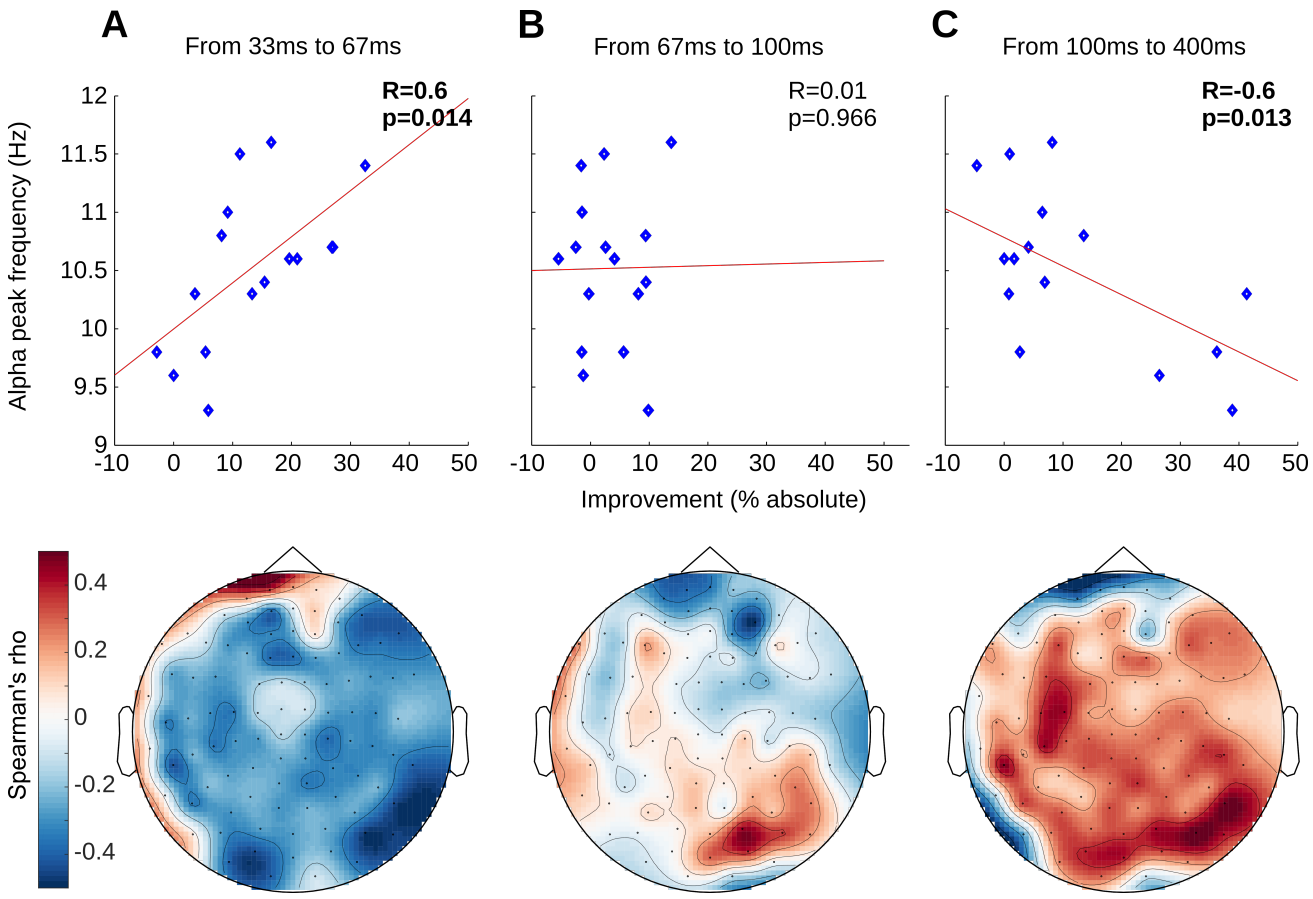

Figure 6. Correlation analysis between performance improvement and resting state alpha peak frequencies. **A.** Top panel: Significant Pearson's correlation between performance improvement from 33ms to 67ms and IAF (R=0.6; p= 0.014). Bottom panel: topographic distribution of Spearman's rho between performance improvement from 33ms to 67ms and IAF. **B.** Top panel: Non-significant Pearson's correlation between performance improvement from 67ms to 100ms and IAF (R=0.01; p= 0.066). Bottom panel: topographic distribution of Spearman's rho between performance improvement from 67ms to 100ms and IAF. **C.** Top panel: Significant Pearson's correlation between performance improvement from 100ms and 400ms and IAF (R=-0.6; p= 0.013). Bottom panel: topographic distribution of Spearman's rho between performance improvement from 100ms to 400ms and IAF. Red lines in top panels A-C show results of linear regression analysis.

As shown in the top panel of Figure 6, we found that individuals with higher resting state alpha frequency showed stronger performance improvement (R=0.6; p= 0.014) in early intervals (33 to 67 ms; Figure 6A). We also found a negative correlation (R=-0.6; p= 0.013) between IAF and performance improvement in late (100-400 ms; Figure 6.C) intervals, but no correlation (R=0.01; p= 0.966) for medium temporal interval (67-100 ms; Figure 6.B). These findings correspond to the general behavioral trend as on average performance improved the least from 67 to 100 ms. On a

descriptive level, the topographies of these correlations showed the strongest correlations between performance differences and resting state IAF in the right occipital sensors. While performance improvement in early intervals corresponded to a negative correlation in the occipital areas (Figure 6.A, bottom panel), performance decline in late intervals corresponded to a positive correlation in these channels (Figure 6.C, bottom panel).

Previous studies have suggested that increased excitability of a relevant sensory region, reflected in pre-stimulus alpha power, could contribute to the perceptual outcome (hits versus misses) of near threshold stimuli. We investigated pre-stimulus alpha power differences between hits and misses for each ISI in sensors corresponding to the visual cortex. We failed to find such statistically significant power modulations ($p>0.05$ corrected for multiple comparison; data not shown).

Recent work has suggested that peak oscillatory frequency varies within an individual during visual perception (Cohen, 2015; Samaha and Postle, 2015; Wutz et al., 2018) and has shown that correctly discriminated visual stimuli exhibit higher alpha frequency than incorrectly discriminated stimuli (Samaha and Postle, 2015). Figure 7 shows that alpha frequency from 344 to 296 ms before stimulus onset was predictive of perceptual outcome ($p= 0.013$, cluster corrected).

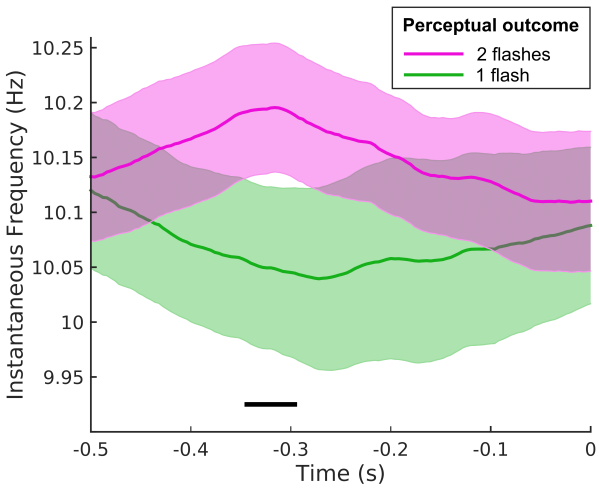

Figure 7. Pre-stimulus instantaneous alpha frequency. Within-subjects analysis of pre-stimulus instantaneous alpha frequency comparing trials in which participants correctly perceived two flashes to those in which only one flash was reported. Significant time points are indicated with a black line ($p<0.05$; permutation test; cluster corrected). Shaded areas depict within-subject standard error.

Discussion

Overall, these findings point to a role of pre-stimulus oscillatory activity in the process of temporally segmenting two visual flashes. First, we tested whether individual participants' IAF was correlated with a performance improvement between increasing subsequent temporal gaps between stimulus flashes (ISI). In the present study, we showed that subjects with higher IAF exhibited the greatest performance improvement at short time intervals when segregating the two flashes is the most difficult. However, an increase of ISI besides the width of a temporal integration window gated by alpha, resulted in no performance differences for individual subjects with high IAF, but in turn was more beneficial for subjects with low IAF. These correlation effects were strongest in posterior sensors, as one would expect in a visual task.

We further found significant perceptual outcome differences (correctly perceiving two flashed versus incorrectly perceiving one flash) in instantaneous alpha frequency in the time before stimulus onset. This finding is in agreement with previous findings (Samaha and Postle, 2015; Wutz et al., 2018). In fact, this result suggests that even though individual alpha frequency can be considered a stable trait, "peaks might actually reflect the average of a small range of frequencies,

within which subtle spontaneous fluctuations are perceptually relevant" (Samaha and Postle, 2015). The present results, underscore the importance of alpha frequency for temporal integration windows.

Because work in perception of near-threshold stimuli suggests that alpha power during the pre-stimulus time interval is a reliable indicator of perceptual outcome, we looked at pre-stimulus alpha power differences between hits and misses for each ISI. If a higher alpha frequency is indicative of a finer temporal resolution, then differences between hits (correctly segregated flashes) and misses (incorrected integrated flashes) should then be reflected by alpha power differences. Specifically, it could be hypothesized that, correctly perceived flash pairs should be preceded by an alpha power decrease in occipital sensors corresponding to primary visual areas. In the present experiment, we failed to find such alpha power differences that were statistically significant. This seems plausible because our experimental set up was of unisensory nature and did not involve spatial or temporal cueing. Weisz et al. (2014) reported contra-lateral pre-stimulus power modulations in a tactile paired-stimulus experiment, but stimulus was always delivered to the left hand along with a temporal cue 500 ms before the stimulus onset. Participants had spatial as well as temporal certainty, which could at least in part account for the strong pre-stimulus power effects reported in their study. Recently published data from our lab (Ronconi et al., 2017) reported alpha phase differences as a reliable outcome indicator, but also did not manage to report significant pre-stimulus power differences between segregated and integrated paired stimuli.

In conclusion, the results of the present study provide evidence for a close link between the speed at which visual information is integrated or segregated and the resting rate of neural oscillations in the brain, with a possible connection being the individual speed of the underlying neural substrate.

CHAPTER 3: THE ROLE OF ATTENTION TO MODALITY IN SPATIAL MULTISENSORY INTEGRATION

Introduction

It has become increasingly clear that top-down control, such as attention or expectation, modulates neural oscillations as well as sensory processing and perception. Ample evidence has accumulated in unisensory processing and perception (Wutz et al., 2014; Frey et al., 2015) and recent studies have confirmed this in the multisensory domain (Gomez-Ramirez et al., 2011; Macaluso et al., 2016; Keil et al., 2017). Neural oscillations associated with attention include local alpha and gamma band power (Jensen and Mazaheri, 2010; Jensen et al., 2007), phase resetting in lower (delta – alpha) frequency bands (Schroeder et al., 2010; Busch et al., 2009) and long-range coherence in beta and gamma band (Gregoriou et al., 2015). In MSI, changes in local alpha band power are reflective of shifting attention in space, towards a sensory modality or a specific stimulus feature. A corpus of experiments found that top-down attention modulates modality-specific low-frequency activity. For example, a shift of attention towards visual stimuli is expressed by alpha-band power decreases over occipital cortex, whereas directing attention to sensorimotor events was also reflected in beta-band power decreases (Bauer et al., 2012). Similar patterns were observed in experimental protocols involving visual-auditory integration (Frey et al., 2014) as well as auditory-somatosensory integration (Leonardelli et al., 2015). Such modulations of low frequencies are thought to reflect a gating mechanism that allows the neuronal population to be ready to process the incoming input (Jensen and Mazaheri, 2010). Others have confirmed that alpha band oscillations take on a suppression mechanism during selective attention (Foxe and Snyder, 2011). Specifically, an increase in alpha power is indicative of lower cortical excitability and higher perceptual threshold, mechanisms crucial for ignoring irrelevant stimuli. Support for this suppression theory comes from

transcranial magnetic stimulation (TMS) studies by Romei and colleagues (Romei et al., 2010). Romei et al. showed that rhythmic 10 Hz occipital and parietal TMS stimulations diminished participants' ability to detect near-threshold visual stimuli contralateral to the TMS stimulation site. Foxe and Snyder further dismissed the notion that alpha band power increase simply reflects a return to baseline levels for areas processing distractors through a series of experiments that found anticipatory alpha to be statistically greater from pre-cue baseline levels (Kelly et al., 2006; Fu et al., 2001). Conclusively, alpha power oscillations modulate the excitatory state of cortical regions reflected through active processing in task-relevant networks, active suppression of task-irrelevant regions or both.

Multisensory integration is not only reflected in pre-stimulus oscillations, but also in evoked oscillations. If MSI truly extends into primary sensory regions, this effect should be reflected by different early evoked responses (ERP and ERF) as well as in differences in evoked gamma band activity. Such activity should then take place at shorter latencies than the later multisensory processing in higher-order cortical areas. As mentioned in Chapter 1, oscillatory neuronal synchronization in the gamma band has been bespoken to play a key role in information coding (Singer, 1999; Gratton, 2018). Following sensory stimulation, two gamma oscillations are usually noted: early evoked gamma oscillations and a late induced gamma oscillation (Başar-Eroglu et al., 1996; Başar, et al., 2001). Early evoked gamma oscillations typically occur within the first 100-150 ms after stimulus onset in cortical areas sensitive to the presented stimulus features, and are timelocked from trial to trial. It is supposed that they result from sensory processing. Early sensory-evoked gamma oscillations have been reported in humans (Tallon-Baudry and Bertrand, 1999): visual-evoked gamma at 25-45 Hz at around 100 ms (Herrmann et al., 2004), auditory-evoked gamma at 30-60 Hz at around 50 ms (Tiitinen et al., 1993) and somatosensory-evoked gamma at 60-95 Hz at around 40-100 ms (Bauer et al., 2006). These sensory-evoked gamma oscillations are sensitive to attentional factors (Fries et al., 2001; Debener et al., 2003; Fell et al., 2003; Bauer et al.,

2006). Conversely, induced gamma oscillations typically occur later at around 240 ms post-stimulus and exhibit variation in latency from trial to trial (Tallon-Baudry & Bertrand, 1999). Therefore, induced gamma oscillations might denote higher cognitive processes (Tallon-Baudry, 2003). Indeed, along with somatosensory-evoked potentials, somatosensory gamma activity has been utilized in neurosurgery in order to identify the hand areas of the primary somatosensory cortex, by electrically stimulating the median nerve at the wrist (Allison et al., 1989; Fukuda et al., 2008). A series of experiments has investigated early multisensory integration effects through evoked responses and early evoked gamma band activity, as gamma frequency oscillations have been associated with selective attention and sensory integration processes.

In an EEG study, McDonald and Ward (2000) showed that auditory capture of visual attention is represented by a negative difference event-related potentials effect. Participants were presented with a visual target that was cued spatially (either valid or invalid) by an auditory tone. When cue and target appeared in close temporal proximity, the negative difference that reflects effects of spatial attention was largest over the occipital cortex contralateral to the target location. This lateralization in the occipital cortex suggests modulation of the early visual cortex by means of spatial attention. In a follow-up study, McDonald and colleagues (2003) furthermore found early activation at 120-240 ms in superior temporal sulcus and gyrus, a brain area known to receive and integrate multisensory information. This activation was accompanied by activation in the fusiform gyrus in the visual cortex at 150-170 ms. Their findings suggest that enhanced visual perception fostered by the crossmodal orienting of spatial attention stems from neural feedback of the multimodal superior temporal cortex to the visual cortex of the ventral processing stream. Senkowski et al. (2007) investigated the effects of different stimulus onset asynchronies in auditory-visual integration while recording EEG data. When the auditory and visual stimuli were presented in close temporal synchrony (0 ± 25 ms), they found early oscillatory gamma band (30-80 Hz) responses at 30-80 ms over medial-frontal brain areas and at 60-120 ms in occipital electrodes. These

multisensory interactions provide evidence for early multimodal interaction effects when multimodal stimuli are presented in temporal synchrony.

Teder-Sälejärvi et al. (2005) investigated spatial constraints in audio-visual integration by presenting unisensory stimuli or bisensory stimuli, either on the right or on the left side, in a target detection paradigm while recording EEG. Participants had to make speeded detection responses to infrequently presented auditory and visual targets or a combination of both within a stimulus sequence. They found behavioral benefits for bimodal stimuli, independent of spatial congruency, compared to either unimodal stimulation, expressed in faster reaction times and higher accuracy. When comparing event-related potentials of the bimodal stimuli with the summed unimodal stimuli, they found interaction effects at 190 ms in electrodes corresponding to the ventral occipito-temporal cortex and at 260 ms in electrodes corresponding to the superior temporal cortical areas. Importantly, they found spatial congruency effects through ERP phase and amplitude modulations in the ventral occipito-temporal cortex at 100-400 ms for visual evoked activity, and amplitude modulations in the superior temporal region later at around 260-280 ms. It is, however, noteworthy that in this study participants only had to respond to targets regardless of modality or spatial location and congruency.

Other studies have investigated early integration effects for auditory and tactile stimulus combinations. For example, Foxe et al. (2000) found by means of high-density EEG recordings early auditory-tactile integration effects at around 65 ms in the postcentral gyrus, specifically in the hand representation area. This was followed by an activation at around 80 ms in the posterior auditory cortices. Sperdin and colleagues (2009) investigated if early auditory-somatosensory integration effects for spatially aligned, as well as misaligned stimulus combinations in auditory association areas contralateral to the side of the somatosensory stimulation (Murray et al., 2005) were correlated to reaction times. When reaction time was relatively fast, they found supra-additive response strength modulations at 40-84 ms post-stimuli in posterior regions of the superior temporal cortex. The term supra-additive refers to the observation that the effect of combined stimulations, in this specific

example bimodal: auditory-somatosensory, was greater than the sum of the response strengths of the constituent unimodal stimulations. Supra-additive modulation in response strength at later times post-stimulus (86-128 ms) exhibited no correlation with reaction times. Importantly, they found no evidence of distinct engaged networks for fast and slow reaction times, but rather for a modulation in their strength.

Work from Quinn et al. (2014) attempted to investigate MSI effect in a visual-tactile detection task by means of intracranial electrocorticography. Stimulation was delivered through a small handheld device and consisted of a brief tactile tap to the thumb or index finger, a brief LED flash at the location of the thumb or index finger, or a temporally synchronous combination of both the visual and tactile stimuli. Observers were instructed to detect and quickly respond to stimuli in the target modality, while ignoring the respective distractor modality. By means of averaged local field potentials and high-gamma-band power (while controlling for nonspecific responses such as stimulus expectancy), they found no evidence of spatial congruency effects, and thus pooled the bimodal stimulation together. Comparing unimodal stimulations with bimodal stimulations, they found MSI to occur in the temporo-parieto-occipital junction (TPOJ) at late latencies after stimulus onset at around 200 ms, but failed to find early temporal and anatomical MSI effects. Their findings are in support of the idea that MSI occurs relatively late and in higher-order multimodal cortices, instead of in primary sensory areas. They point to the supramarginal region, concerned with body image and agency (Blanke et al, 2005). It receives input from auditory, somatosensory and visual modalities (Jones and Powell, 1970). Further evidence comes from reports that lesions to this area can result in multimodal hemi-neglect (Sarri et al., 2006). Taken together, while fMRI studies lack the precise temporal resolution, EEG/MEG studies lack the precise spatial resolution necessary to make irrefutable statements about the temporal aspect and cortical sources of MSI. Consequently, Quinn and colleagues' intracranial findings (bearing in mind that recording sites are small and limited) make a compelling argument for the late-integration model of multisensory processing, at

least for visual-tactile stimulus pairings. Stimuli are initially processed in their respective low-level sensory areas and then converged in higher association areas (Quinn et al., 2014).

Motivation for present study

Multisensory integration has been comprehensively investigated in recent years. However, less work has investigated the behavioral and neurophysiological effects of attention to modality in the context of a multisensory stimulus. In the paragraphs above, we attempted to give an overview of the multisensory links in the spatial attention literature and suggest that the majority of research has focused on attentional switching between spatial locations, which differs from attentional switching between sensory modalities. Moreover, surprisingly few studies exist that focus on investigating the interplay between more than two sensory modalities. This is a crucial step in understanding MSI, as our sensory world is not composed only of bimodal stimuli. Investigating more than two sensory modalities can help to reveal directionality and hierarchy and will enable us to understand more about spatial alignment of different sensory modalities.

One question is how directing attention to two sensory modalities, while simultaneously averting attention from a third distractor modality, is reflected in pre-stimulus alpha/beta frequency power modulations. Furthermore, the majority of multimodal studies discussed here involved attentional switching between spatial locations, and therefore resulted in contralateral power modulations in alpha- and, if applicable, beta-frequency bands. This however is fundamentally different from truly switching attention between sensory modalities without any spatial confounds. The following study aims at investigating the pre-stimulus and early behavioral and evoked effects of multisensory integration and attention to modality by using a trimodal auditory-visual-somatosensory spatial congruency task. On every trial, we presented three discrete, temporally synchronous stimuli in different sensory modalities while recording MEG. Top-down control (attention) was manipulated across separate blocks. Participants were cued to pay attention to various stimulus pairs (audio-

tactile, audio-visual, visual-tactile) and report whether they were spatially congruent (forced choice). We refrained from the use of spatial cueing, and therefore expected no spatial hemispheric attention effects in the pre-stimulus low frequency power bands. In the majority of trials, the respective stimulus combination was congruent, which allowed us to investigate the congruency effects of the unattended stimulus modality. In order to understand if a frequency modulation between the stimulus conditions was to be found in the pre-stimulus period we focused our analysis on the sensory regions corresponding to our stimulus modalities. Furthermore, we investigated if our attentional manipulations caused different early multisensory integration response patterns. If MSI is facilitated purely by endogenous attention (task/goal directed), then one would expect little impact of the unattended sensory modality, whether presented congruent or incongruent, on the attentional modality combination. Previous research findings on the role of attention in MSI have suggested that the role of attention in MSI should be considered as a combination of endogenous and stimulus-driven signaling for the selection of relevant information and behavioral outcomes (Macaluso et al., 2016). Based on sensory dispositions as described in Chapter 1, such that the visual system is more reliable for coding spatial information, bottom-up salience of the visual stimuli could grab participants' attention and thus make the task more difficult, when attention is directed towards the audio-tactile stimulus pair. Furthermore, behavioral benefits should result from trials with complete stimulus location congruency, meaning trials in which all stimulus modalities are spatially congruent.

Materials and Methods

Participants

A group of 21 healthy volunteers with no known neurological or psychiatric disorders took part in the experiment. All volunteers had normal to corrected-to-normal vision and reported a normal sense of hearing and touch. Of those, two subjects had to be excluded from the analysis, one due to excessive external artifacts in the MEG data and one due to unavailable HPI coil measurements in 2

experimental runs. Three additional subjects were excluded due to an excessive amount of incorrect trials (>50 %) in at least one condition, resulting in 16 subjects (8 female, mean age: 26.8 years, SD: 3.7 years, all right handed) to be included in the data analysis. All participants gave written informed consent before the experimental session and received monetary reimbursement for taking part in the study. The study had prior approval by the Ethical Committee of the University of Trento and was conducted in accordance with the Declaration of Helsinki as revised in October 2008. The approximate duration of the entire study, including preparation, was 90 minutes.

Stimuli

Auditory stimuli consisted of a 100 ms long tone (44100 Hz sample rate, 16 bits resolution) presented through tubal insert MEG-compatible headphones (SOUNDPixx, VPixx Technologies Inc., Saint-Bruno, QC, Canada) to either the right or the left ear of the participant. Visual stimuli consisted of a Gaussian-enveloped luminance patch of approximately 1° of visual angle diameter at maximum contrast on a uniform grey background and was back-projected by a DLP projector (Panasonic PT-D7700E) on a translucent screen at a refresh rate of 120 Hz. The stimuli were presented for 100 ms at 4° visual angle horizontally either to the right or left of the screen center.

Tactile stimuli were a 100 ms stimulation delivered to the tip of the right or left middle finger, using two finger modules of a piezo-electric stimulator (Quaerosys, Schotten, Germany) with 2 x 4 protruding rods. The stimulators were attached to the fingers with tape. The participant's hands were cushioned to prevent any unintended pressure on the module (Frey et al., 2016) and placed at approximately shoulder width to be aligned (along the body axis) with the visual and auditory stimuli.

The precise timing of the visual, auditory and tactile stimulation was calibrated via a photo diode placed at the upper left corner of the projection screen and an oscilloscope and controlled with a data and video processing peripheral (DATAPixx, VPixx Technologies Inc., Saint-Bruno, QC, Canada).

Task and design

Participants performed six blocks of a spatial localization task. Each block consisted of 120 trials in which participants were to judge if a given stimulus pair (AV, AT, VT) was spatially congruent (yes/no). All three stimuli — auditory (A), visual (V) and tactile (T) — were presented temporally synchronous during each trial. The order of blocks was randomized for each participant.

At the beginning of each block, participants were instructed to pay attention to two of the three stimuli: audio-visual (AV), audio-tactile (AT) or visual-tactile (VT) and judge if these were spatially congruent (both on either the same right or left side). Responses were recorded via forced choice (yes/no) using a MEG-compatible response box (ResponsePixx) with the right or left index finger (counterbalanced between participants). In the present study, we focused on multisensory integration. We were specifically interested in the effects of attention to modality and spatial stimulus congruency. Therefore, in 70-80 % (randomly distributed) of the trials, the respective stimulus combination (AV, AT or VT) was congruent (meaning "yes" they were aligned) while in the remainder of the trials they were not ("no"). The 3rd stimulus modality (for example if paying attention to AV: T) could then equiprobably either act as a distractor (spatially incongruent to the stimulus pair) or an enhancer (congruent, on the same side as the stimulus pair). Figure 8 illustrates a trial sequence: Each trial started with a central fixation cross (black, 1° visual angle on a uniformly grey background for a variable pre-stimulus interval (1-2 s, randomly distributed), followed by an experimental trial. During the inter-trial interval (ISI), the fixation cross changed color to red (wrong) or green (correct) based on participant response. The fixation cross changed back to black (0.7-1.3 s, randomly distributed) before the onset of a new trial. Participants were instructed to maintain fixation on the central fixation cross throughout the experimental trials in order to minimize eye movement artifacts. Stimuli were presented to subjects in a dimly lit magnetically shielded room (AK3B, Vakuumschmelze, Hanau, Germany). The stimuli were generated on an HP Intel Quad core

computer using Matlab 8.0 (MathWorks) and Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997).

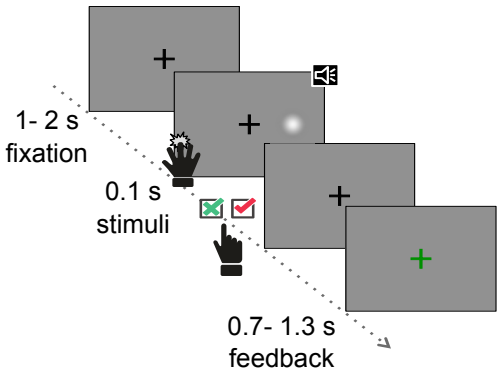

Figure 8. Illustration of a trisensory experimental trial. After a variable inter-trial interval during which participants fixated on a central fixation cross, all three stimuli were presented for 100 ms. Participants were to indicate if the stimulus pair they were paying attention to was spatially aligned by pressing one of two buttons ("yes" or "no").

Data acquisition

MEG data were recorded at a sampling rate of 10 kHz using a 306-channel (204 first-order planar gradiometers, 102 magnetometers) VectorView MEG system (Elekta-Neuromag Ltd., Helsinki, Finland) in a magnetically shielded room (AK3B, Vakuumschmelze, Hanau, Germany). Hardware filters were adjusted to band-pass the MEG signal in the frequency range of 0.01-1,000 Hz. Prior to the experimental runs, a subject-specific head-frame coordinate reference was defined, in order to localize the head position of the subject within the MEG helmet. The cardinal points of the head (nasion, left and right pre-auricular points), the location of five head-position indicator (HPI) coils, and a minimum of additional 200 head-shape samples were digitized for motion tracking (3Space Fastrack; Polhemus, Colchester, VT) at the start of each session. The subject's head position relative to the HPI coils and the MEG sensors was estimated before each experimental run to ensure that no large movements occurred during the data-acquisition procedure.

Data preprocessing

Environmental noises (noise created by the piezo-electric tactile stimulator and other external inferences) were removed and the data was co-registered in order to remove small head movements across the six separate measurement runs through Signal Space Separation with spatio-temporal extension (tSSS; Taulu & Hari, 2009) implemented through the MaxFilter software version 2.2.15 (Elektra-Neuromag Ltd., Helsinki, Finland). Prior to that, data was visually inspected and noisy channels were excluded from the tSSS maxfiltering. The total translation across session ranged from 0.4-15.2 mm across participants (median = 3.14 mm). The rank of the data covariance after pre-processing with MaxFilter varied between 68-71 (median = 69). Data were then analyzed using the Fieldtrip toolbox (Oostenveld et al., 2010), the CoSMoMVPA toolbox (Oosterhof et al., 2016) and custom-built Matlab functions in combination with Matlab 8.1 (MathWorks, Natick, MA). Residual artifacts induced by the piezo-electric stimulation (see Buchholz et al., 2011) as well as residual cardiac signals were removed using Independent Component Analysis (2-4 components per subject).

As we were interested in the integration of various stimulus pairs, epochs of 4 s (2.5 s pre and 1.5 s post) were centered on the stimuli onset. Zero seconds therefore represent the simultaneous onset of all three stimuli. Trials were visually inspected for additional possible artifacts and contaminated trials were excluded from further processing. Across subjects, an average of 3.06 % (SD: 1.74 %) of the 720 trials were rejected, which, depending on the subject, resulted in 659 -710 artifact-free trials for further processing.

As maxfiltered and concatenated data results in an ill-conditioned estimate of covariance, the number of components was reduced to that accounting for 99 % of variance in the covariance matrix based on the components eigenvalues for each participant (59-63 number of components, median = 61).

Data Analysis

Time-frequency analysis in sensor space

For each participant, time-frequency analysis was performed, averaged over correct trials (number of trials per condition: AV: mean= 216.5625, SD= 14.5601; AT: mean= 206, SD= 17.705, VT: mean= 211.3125, SD= 21.9385) between 2-30 Hz (in steps of 1 Hz). An adaptive sliding window with a length (Δt) of 5 cycles of the respective frequency was applied and shifted in steps of 10 ms between –1.0 to 0.5 s. A Hanning taper was applied, yielding a spectral smoothing of $1/\Delta t$.

Since our interest focused on sensory regions, we further investigated sensors of interest (SOI) in the respective regions. Data were analyzed in three regions of interest by selecting and averaging over sensors over the right and left temporal areas corresponding to the auditory cortex, bilateral parietal areas corresponding to the somatosensory cortex, as well as right and left occipital areas corresponding to the visual cortex.

Additionally, high-frequency power (30-120 Hz in steps of 2 Hz) was estimated using multitapers with a fixed time-window length of 200 ms and a fixed smoothing factor of 10 Hz for correct trial in each condition separately for each stimuli congruency and hemisphere. Activity was then baseline normalized using an interval of – 500 to 0 ms before stimulus onset.

Event-related fields (ERFs) analysis in sensor space

Data for each participant was segmented from 500 ms before stimulus onset to 1000 ms after stimulus onset and 40 Hz lowpass-filtered. Then the event-related fields (ERF) were computed only on correct trials and congruent trials. Additionally we computed the ERFs for each condition as a function of congruency (ignored modality congruent/incongruent with attended modality pairing) separately for each hemisphere. The ERFs were baseline normalized using an interval of -500 to 0 ms before the stimulus onset.

Statistical testing in sensor space

On sensor level, statistical significance was assessed by means of cluster-corrected z-statistics of differences between attentional conditions using a threshold-free method for clustering (tfce) and a Monte Carlo permutation approach for multiple-comparison correction as implemented in CoSMoMVPA (Oosterhof, et al., 2016). This type of test was used to control for the multiple-comparison problem (type I error). A tfce-based one-sample t-statistics against zero with Monte-Carlo randomization was computed for each attentional condition contrast on the pre-stimulus period (-1 to 0 s) for different frequency bands (alpha: 8-15 Hz, beta: 16-30 Hz) separately for the two sensor types. The same statistical testing was performed on the SOI (sensors of interest) comparing attentional conditions as well as contrasting attending versus ignoring each sensory modality in the respective SOI regions, correcting for multiple comparison in time and frequency.

In the post-stimulus time period, we performed equivalent tfce-based one-sample t-statistics as described above for the SOI separately in each hemisphere in the low (40-60 Hz) and high (60-90 Hz) gamma band in the time period from stimulus onset (0) to 150 ms. With the same parameters, we also tested if attentional condition and unattended stimulus congruency resulted in early gamma power modulations in the constituent sensory regions in each hemisphere by means of a repeated measure ANOVA (corrected for multiple comparison).

For statistical analysis of early ERF differences between the attentional conditions, we calculated the average ERF peak amplitude of all three stimulus conditions (AT, AV, VT) for each of the three sensory regions of interest (occipital, auditory and somatosensory channels) from stimulus onset to 200 ms separately for each sensor type. Statistical testing by mean of ANOVA F-statistics with Bonferroni-corrected post hoc t-tests was then computed between the three stimuli conditions for a time window of 20 ms centered on the peak amplitude computed in the previous step.

Furthermore, we investigated the entire early time course (0-200 ms) in order to understand if stimuli congruency and attentional condition resulted in different ERF signatures in each of the sensory

regions. A repeated-measure ANOVA with correction for multiple comparisons (CoSMoMVPA) was calculated for each sensory region of interest in each hemisphere separately, comparing each relevant attentional stimuli condition as well as the congruency of the irrelevant stimulus condition (congruent or incongruent along with attentional stimulus pair).

Results

Behavior

Accuracy was measured as the proportion of correct responses (yes or no) of total trials in each attention condition. Across participants (N=16) mean accuracy was 92.45 ± 1.41 % (SD: 5.63%) for AV trials, 89.06 ± 1.86 % (SD: 7.45%) for AT trials and 90.90 ± 2.23 % (SD: 8.94 %) for VT trials. Although mean accuracy was slightly higher in the AV condition, there was no significant difference between conditions ($F(2, 30) = 2.291$, $p= 0.119$).

A repeated measure ANOVA determined that for correct responses, mean reaction times (RTs) differed statistically significantly between attention conditions ($F(2, 30) = 5.550$, $p= 0.009$). Post hoc tests using Bonferroni correction revealed that participants were slightly faster in VT trials than in AV trials (889.02 ± 32.42ms vs. 902.65 ± 24.24ms respectively), which was not statistically significant ($p= 0.605$). However, responses to AT trials (972.31 ± 28.15 ms) were significantly slower than AV ($p= 0.005$) as well as VT trials ($p= 0.021$).

As the focus of the study was on multisensory integration, we next examined trials with spatially congruent stimuli in more detail. Specifically, we had hypothesized that a complete stimulus location congruency (meaning the irrelevant, unattended stimulus modality also being spatially congruent along with the attentional modality pair) would be manifested in behavioral benefits in comparison to trials in which it was not. For a detailed overview of accuracy and reaction

times for each spatially congruent attentional condition, split by ignored stimulus modality, see Table 1 at the end of this section.

Accuracy was highest in AV trials (95.807 ± 0.922 %), followed by VT (93.898 ± 1.545 %) and AT trials (93.25 ± 1.412 %). A two-way repeated-measure ANOVA (alpha set at .05) on the effects of attention condition (AV, VT, AT) and stimuli congruency (ignored modality spatially incongruent or congruent) showed no main effect of condition, F ((2, 30) = 2.109, p= 0.139. Importantly, the main effect of stimuli congruency, and the interaction between the effects of attention condition and stimulus congruency were both significant, F (1, 15) = 29.587, p<0.0005 and F (2, 30) = 5.638, p= 0.008, respectively. The effect of stimulus congruency was further examined, and revealed that on average, participants' accuracy was 4.72 ± 0.87 % higher when the irrelevant stimulus also was spatially congruent with the attentional modality pair. Furthermore, simple main effects analysis showed that accuracy was significantly higher in all three attention conditions (AT: p= 0.001, AV: p= 0.001, VT: p= 0.005) when all three stimuli modalities were spatially congruent (see also Table 1 for mean behavioral results in detail).

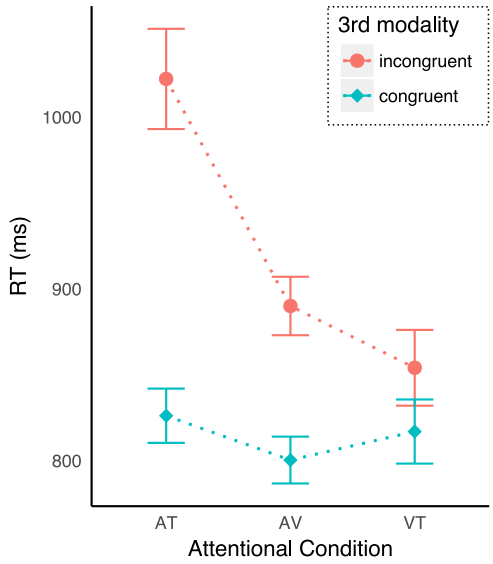

Figure 9. RTs as a function of attention condition and stimuli congruency. Error bars denote within-subject standard error.

Figure 9 presents the mean reaction times in correct responses as a function of attention condition (auditory-tactile, auditory-visual, visual-tactile) and stimuli congruency (ignored modality spatially incongruent or congruent). These means were subjected to a two-way repeated-measure ANOVA with the same factors and alpha set at .05. Overall, RTs showed significant differences in attention conditions $F(2, 30) = 6.874$, $p = 0.003$. Post hoc testing using Bonferroni correction revealed that participants responded significantly slower to AT trials (924.192 ± 27.558 ms) than AV trials (845.2 ± 22.796 ms; $p = 0.009$) as well as VT trials (835.523 ± 30.962 ms; $p = 0.042$). Although participants responded slightly faster in VT trials than in AV trials, this difference was not statistically significant, $p>0.05$). The main effect of stimulus congruency was significant, $F(1, 15) = 57.751$, $p< 0.0005$, showing that RTs in trials where all three stimulus modalities were spatially congruent were 107.619 ± 14.161 ms shorter than trials in which the irrelevant stimulus modality was spatially incongruent with the attentional modality pair. Importantly the ANOVA revealed a significant interaction between attention condition and stimuli congruency, $F(2, 30) = 19.772$, $p<0.0005$. This interaction was further examined in detail by simple main effects analysis. Simple main effects analysis showed that mean RTs were significantly shorter when all three stimuli modalities were spatially congruent compared to when only the attentional pair was spatially congruent in the AT ($p <0.0005$, mean difference 196.07 ± 26.9 ms), the AV ($p = 0.0005$, mean difference 89.69 ± 17.69 ms) and in the VT ($p = 0.027$, mean difference 37.11 ± 15.12 ms) condition.

Table 1. Behavioral results of accuracy and mean reaction times for correct responses with standard errors in the three different stimulus conditions split by total congruency. Auditory-tactile attention pair with visual spatially incongruent (AT2) and congruent (AT3). Auditory-visual attention pair with tactile spatially incongruent (AV2) and congruent (AV3). Visual-tactile attention pair with auditory spatially incongruent (VT2) and congruent (VT3).

	AT2	*AT3*	*AV2*	*AV3*	*VT2*	*VT3*
Accuracy	89.039 %	97.468 %	94.508 %	97.107 %	92.335 %	95.462 %
S.E. Acc.	2.335 %	0.916 %	1.101 %	0.838 %	1.744 %	1.479 %
Mean RT	1022.229 ms	826.155 ms	890.038 ms	800.363 ms	854.077 ms	816.969 ms
S.E. RT	38.966 ms	19.035 ms	25.625 ms	23.220 ms	35.689 ms	27. 530 ms

Pre-stimulus time frequency

In order to investigate if top-down control (attention) modulated pre-stimulus brain oscillations, we looked at each of the three attentional conditions separately, with comparison to the other two attentional conditions. In other words, we compared time frequency responses of trials in which participants had to pay attention to a certain modality (+) with trials in which they ignored that same modality (-) in the alpha (8-15 Hz) and beta (16-30 Hz) band in the time period prior to stimulus onset (-1 to 0 s). Figure 10 depicts these results. As we had hypothesized to find pre-stimulus power modulations in these lower frequency bands in the corresponding sensory cortices, we restricted the statistical comparison described above for modality attended (+/-) to the respective sensory regions (Figure 10). In the statistical comparison for Visual +/- (attend vision vs. ignore vision: mean (AV+VT) vs. AT) a significant negative cluster was found in the combined gradiometers corresponding to occipital regions at around -400 ms to stimulus onset in the alpha band from 10-15 Hz (z_{min} = -2.737; corrected for multiple comparison in frequency and time). Additionally, we found significant clusters in the combined gradiometers corresponding to auditory regions for Auditory +/- in the beta band between 24 and 30 Hz at around -680 to -620 ms, -150 to -100 ms and -20 to 0 ms prior to stimulus onset (z_{min} = -2.209). In an exploratory investigation at the whole brain level, no comparison of attended modality survived correction for multiple comparisons.

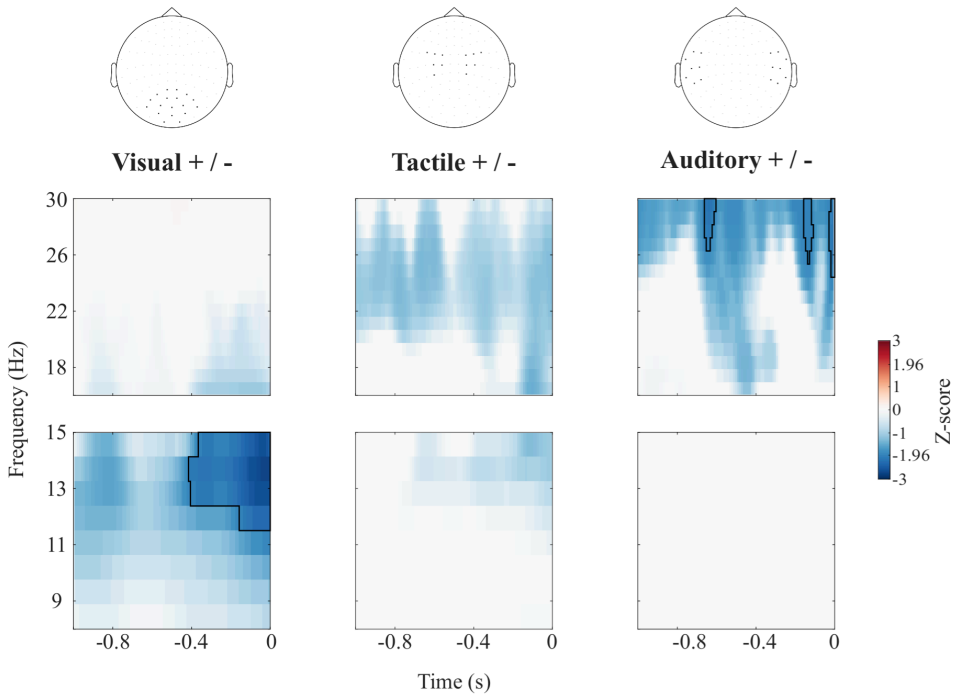

Figure 10. Pre-stimulus alpha and beta time-frequency contrasts for attend to versus ignore sensory modality in sensors of interest for sensory areas (combined gradiometer). Marked with black outlines are the statistically significant time-frequency windows ($p<0.05$ corrected).

Following up on this finding, we further investigated if multisensory attention modulated pre-stimulus brain oscillations in the three sensory regions by comparing time frequency responses between the attentional conditions in the alpha and beta band prior to stimulus onset (same parameters as above). Figure 11 illustrates the results of these statistical comparisons. Statistical comparison of AV and AT revealed a negative significant difference in the alpha band between 10 and 15 Hz from -400 ms to stimulus onset ($z_{min} = -2.334$) in occipital sensors, as well as a positive significant difference in somatosensory sensors in the beta band from 16 to 25 Hz from -380 to -300 ms ($z_{max} = 1.99$). Statistical comparison of AT and VT trials revealed a negative difference in the beta band from 20.5-30 Hz in auditory sensors ($z_{min} = -3.719$) from -100 ms to stimulus onset. Statistical comparison of VT and AV revealed no statistically significant differences.

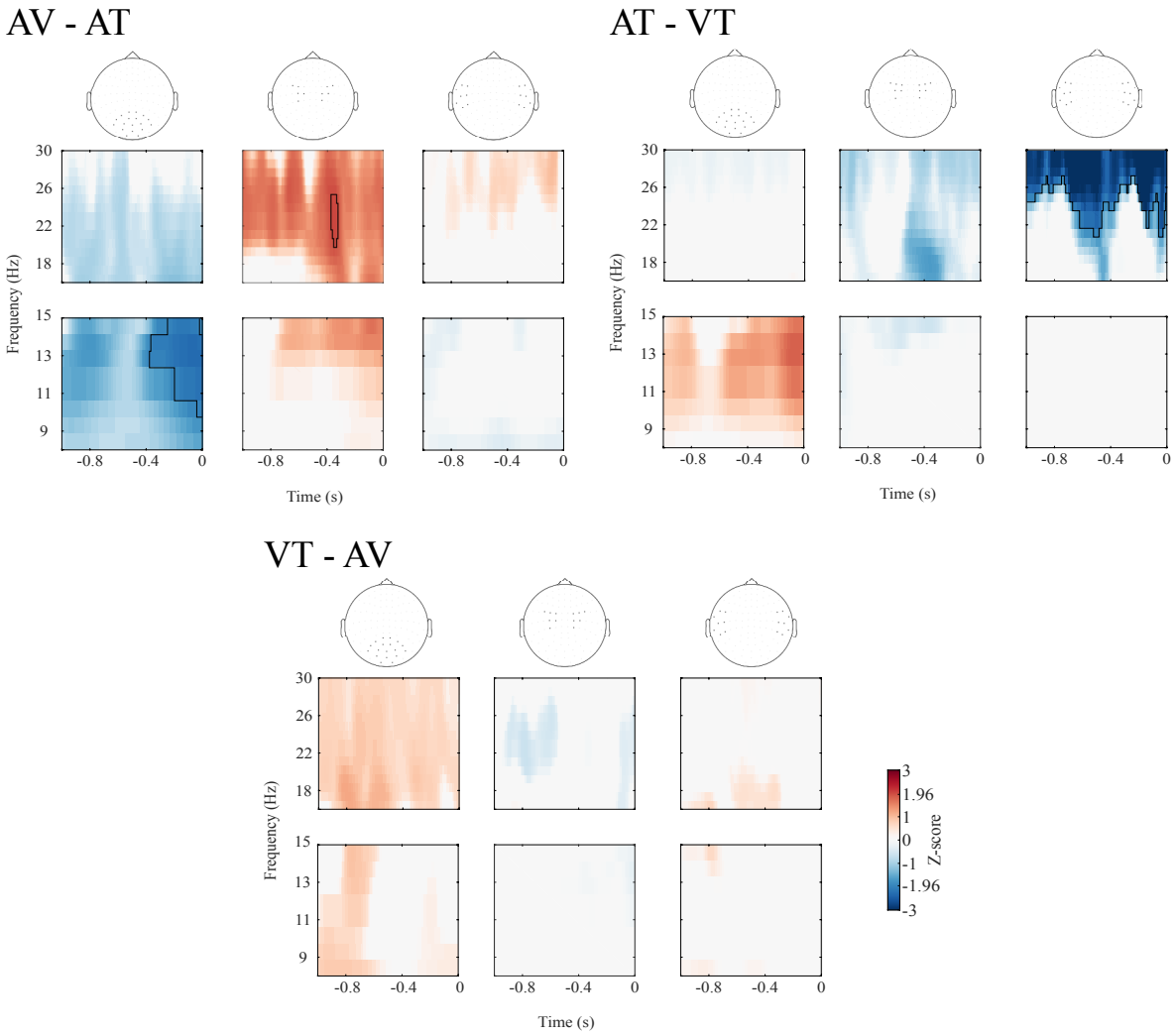

Figure 11. Pre-stimulus alpha and beta time-frequency contrasts for all different experimental attention conditions in sensors of interest for the three sensory areas (combined gradiometer). Marked with black outlines are the statistically significant time-frequency windows ($p<0.05$, corrected).

Post-stimulus event-related neural activity

Next, we investigated evoked responses (ERFs). Despite the fact that the stimulus was identical, we hypothesized that our attention manipulation would cause a different evoked response pattern. First, we looked at the three conditions (VT, AV, AT) in each of the three sensory regions from stimulus onset to 200 ms separately in order to investigate early sensory processing and integration (Figure 12). Statistical comparisons of peak amplitudes of each condition in combined gradiometers corresponding to the occipital cortex around 113 ms (103-123 ms) revealed a statistically significant difference, $F(2, 14) = 14.0725$, $p= 0.0097$. Post hoc (Bonferroni corrected) tests revealed those differences to be between AV and VT ($p= 0.0052$; Bonferroni corrected). In selected combined

gradiometers corresponding to the somatosensory cortex, we failed to find statistically significant differences between each condition's peak amplitude at mean amplitude time (103-123 ms) (F (2, 14) = 3.1034, p= 0.2681. In combined gradiometers corresponding to the auditory cortex we found a statistically significant difference between peak amplitudes at mean amplitude time in those sensors (98-118 ms), F (2, 14) = 10, 8132, p= 0.0224, with the differences being between AT and AV (p= 0.0104, Bonferroni corrected). Statistical tests in the respective selected magnetometers failed to reach significance in all conditional comparisons (data not shown; p>0.05).

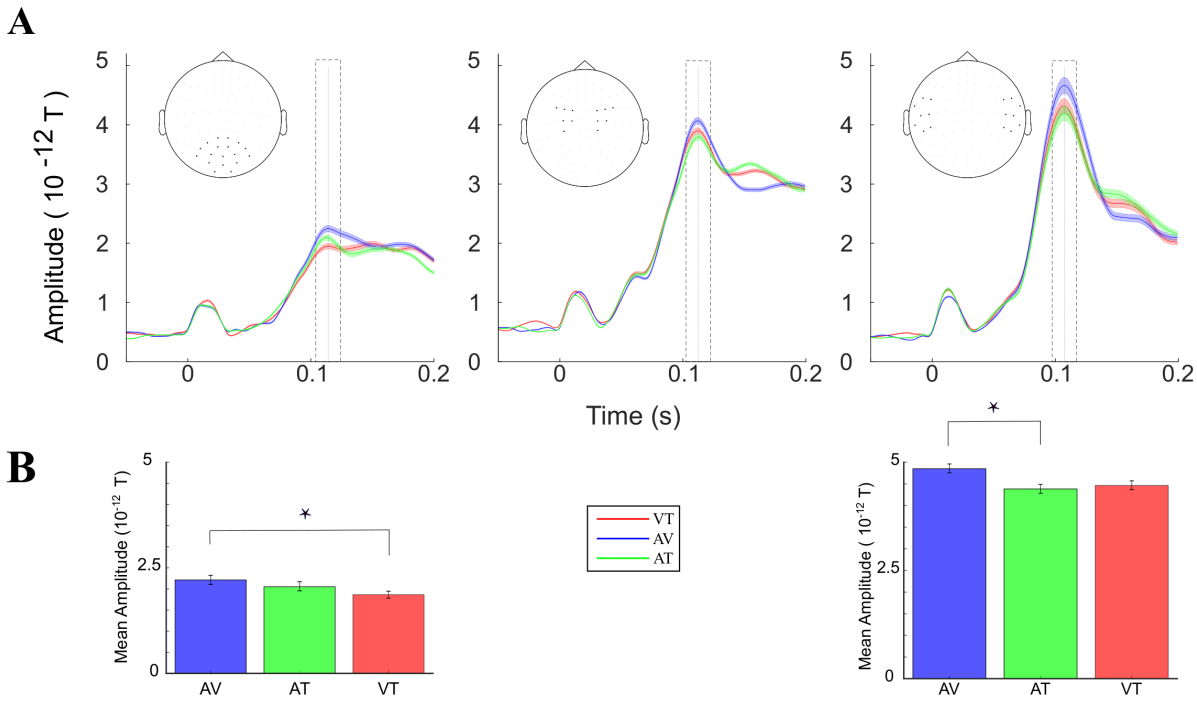

Figure 12. Sensor-level global field power. **A**. Sensor-level event-related global field power in visual-tactile (red), audio-visual (blue) and audio-tactile (green) attentional conditions for sensors of interest in combined gradiometer data. Marked with dashed rectangles: 20 ms time window centered on overall mean peak amplitude in respective averaged sensors of interest. **B**. Post hoc results of attention condition mean average at mean peak amplitude time window in sensors of interest.

Following up on these analyses of the effects of attention condition on the ERFs, we next took a closer look at stimuli congruency and attentional conditions in early ERFs (0-200 ms) of combined gradiometer to scrutinize if there are early multisensory integration effects in the sensory regions. As the attended stimuli could appear on the right as well as the left side, but involved no spatial cueing,

we investigated ERFs separately for each hemisphere (Figure 13). In sensors over right as well as left auditory cortex (Figure 13.C), repeated-measure ANOVAs between AV and AT trials showed a statistically significant main effect of condition from 93 to 112 ms (z = 2.4229) and 99-118 ms, respectively (z = 2.1781). Furthermore, we found a significant interaction effect in sensors over right auditory cortex (Figure 13.C top) from 25 to 32 ms (z = 2.3417) and 107-113ms (z = 2.3418). In sensors over left somatosensory cortex (Figure 13.B bottom), a repeated-measure ANOVA between AT and VT revealed a statistically significant main effect of congruency between 156 and 161 ms (z = 1.9936).

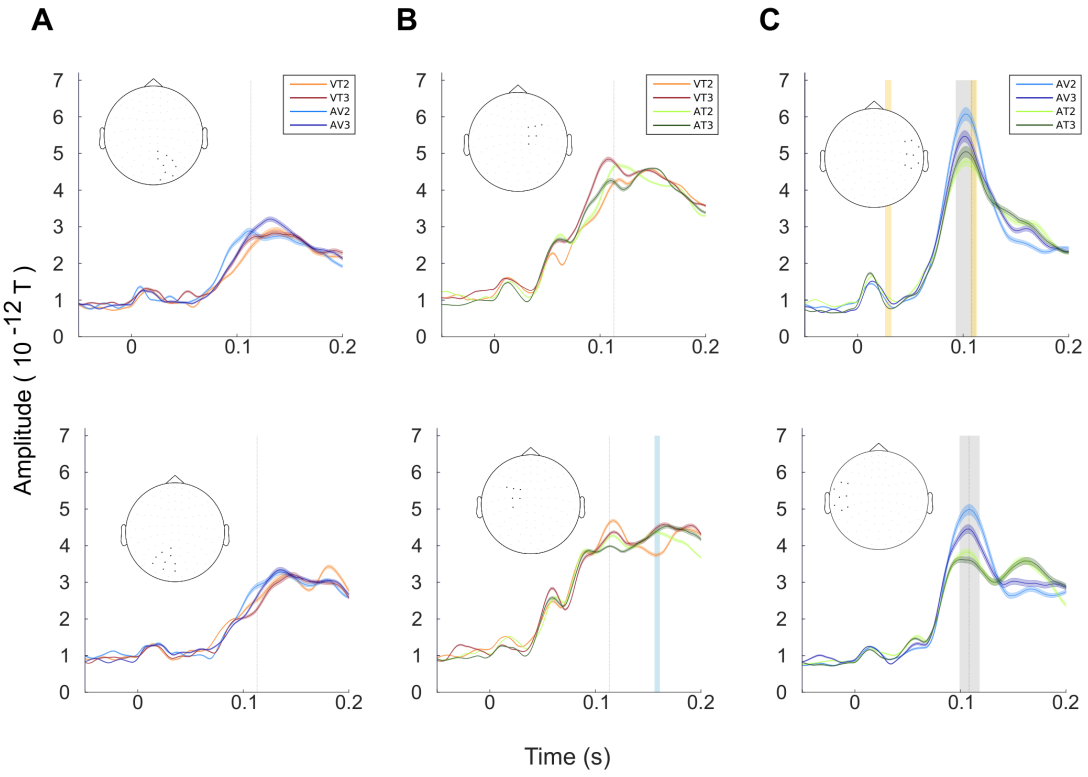

Figure 13. Sensor-level global field power. Sensor-level event-related global field power as a function of attention condition and stimulus congruency in right (top) and left (bottom) hemifield for sensors of interest over the respective sensory areas in combined gradiometer data (**A**: visual, **B**: tactile **C**: auditory). Marked with dashed line: overall mean peak amplitude in respective averaged sensors of interest (see Figure 12A). Significant main effect of attention condition (grey shading, AV>AT in auditory cortex (Panel C); $p<0.05$ corrected), significant main effects of stimulus alignment (light blue shading, unattended modality incongruent> unattended modality congruent in left somatosensory cortex (Panel B); $p<0.05$ corrected) and significant interaction effect (yellow shading in right auditory cortex (Panel C); $p<0.05$ corrected) of ERFs observed in sensory regions (visual, somatosensory and auditory cortices).

Post-stimulus time- frequency

Following up on the ERF results, we looked for early multisensory integration effects in the three sensory regions by means of gamma band power. Gamma, especially high gamma power, has been found to be a relatively direct measure of local high-frequency synaptic and spiking activity (Quinn et al., 2014). First, we looked in each of the three sensory regions separately for each hemisphere to contrast attending or ignoring the corresponding sensory modality (Visual +/-, Auditory +/-, Tactile +/-). We failed to find any statistically significant difference at these short latencies in the low or high gamma band that would support early integration effects in primary sensory regions. Following up on the repeated-measure ANOVA results found in the time domain, we calculated repeated-measure ANOVAs in the low and higher gamma band power. Again, we failed to find statistically significant effects of attentional condition and congruency, as well as interaction effects in each sensory region at short latencies of 0 to 150 ms (data not shown).

Discussion

The goal of the present study was to investigate the role of attention to modality in multisensory integration. To this end, a spatial congruency task was used, in which we manipulated multisensory attention, while keeping the stimuli constant.

Both attention to sensory modality combinations and spatial congruency effects were reflected in behavioral outcomes, here measured in accuracy and reaction times. Paying attention to a multisensory combination that includes the visual modality elicited a significant reduction in reaction time. In contrast, we found longer reaction times in trials where participants had to pay attention to auditory-tactile stimuli (hence, ignoring vision). There were statistically significant behavioral benefits, increases in accuracy and reductions of reaction times, when the ignored, irrelevant stimulus modality was also congruent along with the attentional modality pair, and could even be described as an enhancer. This effect, clearly visible in Figure 9, was strongest in the audio-tactile

attention condition, where vision took on the role of the enhancer. In other words, ignoring the visual input, when attention is directed away from it, seems to be more difficult than ignoring other sensory modalities. One possible explanation could be that ignoring the irrelevant visual stimuli requires additional processing steps in mapping the relevant stimuli in the right/left space. This would be in line with the idea that much of the human multisensory experience is guided by vision. This seems plausible on the grounds of the higher spatial resolution of the visual system in comparison with other sensory modalities.

Our experimental manipulations resulted in alpha power modulations in the time before the stimulus onset in corresponding sensory areas. As expected, alpha power differences were most pronounced over the occipital cortex when attention was directed towards sensory combinations involving vision versus ignoring vision. Importantly, there was no spatial cueing involved in our task. Rather than shifting attention between hemifields, our participants had to attend to the space as a whole. Furthermore, the onset of the stimuli was jittered, in order to avoid temporal expectancy effects reflected in pre-stimulus alpha and beta power (van Ede et al., 2011; Bauer et al., 2014). Here we show that, even with spatial and temporal uncertainty, attending to a specific sensory modality influences pre-stimulus oscillatory power in the alpha and lower beta band: although not all statistical comparisons reached significance, they clearly reflect a trend. This is in line with the concept of alpha power modulations as a function of gating through inhibition (Jensen and Mazaheri, 2010). In the time before the arrival of the stimulus, sensory areas that were task-relevant showed a decrease in alpha (and beta) power, whereas task-irrelevant areas were inhibited, demonstrated by an increase in alpha power. Whereas spatial orienting of attention has repeatedly been shown to involve contralateral modulations of alpha and beta oscillations within sensory cortices, we here provide a novel approach that suggests that orienting attention to an upcoming sensory event involves alpha and beta power modulations that are not spatially specific. Although these power differences were

not as pronounced as in experiments that involved spatially orienting attention, they were present in all of our comparisons.

Despite the identical stimuli, our experimental modulations revealed early MSI effects at around 100 ms in sensory regions. Specifically, we found statistically significant differences in ERFs in visual areas between the two experimental conditions that involved attending to vision (AV and VT), and similar results in auditory areas (AT and AV). Interestingly, we found that the audio-visual stimulus pair elicited the strongest evoked response. Sperdin and colleagues (2009) previously found that the strength of the early evoked response in an audio-tactile integration was correlated to faster reaction times. In the present experiment, we found the fastest reaction times when attention was directed to visual-tactile and audio-visual stimuli.

Previous studies have shown that gamma frequency oscillations are associated with selective attention and sensory integration and it can therefore be hypothesized that gamma activity synchrony may serve as the basis of cerebral functionality and cortical communication (Başar et al., 2001). For low and high gamma power, we did not observe any statistically significant early integrative attention effects in the three sensory cortices from our trimodal stimulations. Although evoked fields and (high) gamma power are often reported together, they show different temporal-spatial patterns and carry out different computational mechanisms (Quinn et al., 2014). This lack of significant gamma power effect might not reflect the actual neuronal reality. It is well known that outside of intracranial recordings, gamma power differences can be difficult to capture (Nunez and Srinivasan, 2010). Instead of making inferences that, the lack of significant gamma power differences in our MEG data suggests support against early sensory integration effects in a temporal and anatomical sense, we conclude that further investigations that include sensitive measures are necessary. It is important to point out that although MEG data has high temporal resolution, it fails to provide high spatial resolution. Therefore, our findings are to be interpreted with some cautions with respect to specific underlying anatomical sources.

The spatiotemporal evoked response field profile of multisensory stimuli activity in the current study was supportive of early integration effects of multisensory processing. Taken together with results from other multisensory studies, our findings could be interpreted as supportive of a mixed model of multisensory processing. This model postulates that early unisensory processing can be modulated by other sensory inputs, and later sensory streams can be integrated into spatially precise higher-order multisensory representation. Important here is the fact that we found different integrative effects in different sensory cortices. While we found evidence via evoked responses for early multisensory integration in the visual as well as auditory cortex, ERFs in the somatosensory cortex exhibited no such pattern. This is in line with findings by Quinn et al. (2014) who failed to find early multisensory integration effects for a simple visual-tactile detection task in the somatosensory cortex. However, importantly unlike Quinn et al. (2014), we did find evidence for integrative effects in the visual cortex. In our study, the task was more demanding, as it required participants to pay attention to two sensory modalities simultaneously, while ignoring a third, such that attentional demands and task difficulty could be reflected in the present findings of our study compared to others.

One potential concern of our study is that although stimulations in the auditory, tactile and visual domain occurred temporally synchronous, they did not come from the exact same spatial location, but were rather spatially aligned along a diagonal in space. In particular, while the auditory stimulus as well as the tactile stimulus occurred in participants' personal body space (through in-ear headphone or on the fingertips), the visual stimulus occurred rather in space, not body-centered (on a screen in front of participants). In that sense, further investigations are necessary in order to gain a bona fide understanding of MSI. We propose a stimuli design in which the stimuli from all sensory modalities is delivered through the same device, similar to a mobile phone that rings, vibrates and flashes in order to attract attention to a call. Such experimental set up would facilitate an integration of input from different sensory modalities from the same location. An additional independent

presentation of stimuli from the different sensory modalities would further enable us to compare unisensory stimulations to multisensory stimulations, controlling for nonspecific effects (see Quinn et al., 2014). An interesting further extension of such an experimental setup could include conflicting somatotopic reference frames by having participants cross their hands while touching the stimulus device (similar to that in Noel and Wallace, 2016).

In conclusion, the results of the current study provide evidence that top-down control, evident in cortical excitability in the respective appropriate sensory regions, facilitates multisensory integration. While some studies of MSI argue that integration only occurs at anatomically and temporally late stages, in the present study we were able to demonstrate early multisensory effects of attention to modality in sensory regions. We would argue that this discrepancy speaks for a mixed model of multisensory processing, as well as for further analyses that go beyond sensor level effects, taking source space and connectivity measures into account.

CHAPTER 4: IS VISUAL ATTENTION RHYTHMIC?

Introduction

When dealing with a dynamic environment, visual cognition can be described as the segmentation of an environment into elements in space and sampled rhythmically in time. Selective visual attention is a crucial feature of visual cognition that enables the observer to focus on certain specific elements, while filtering out irrelevant information of an environment. These attended elements are processed in greater detail but are also subject to resource constraints, as the visual system is limited to represent only a handful objects at once. This is explained by the limited processing theory, which states that the visual system is able to individuate and keep track of approximately five visual elements (Pylyshyn, 2001). One suitable, classic experimental setup that has been used to demonstrate this is multiple object tracking (MOT). However, MOT is a task we as humans also observe in our everyday life. For instance, while driving in traffic, we are aware of the locations of other cars around us, or when playing team sports.

Multiple object tracking (MOT) is an experimental paradigm that has lent itself at various times to researchers' attempts to uncover the workings of visual attention. An attempt by the Yale Perception & Cognition lab to compile a list of all published MOT studies currently tallies 173 papers in 48 journals (Scholl, 2017). Introduced in 1988 by Pylyshyn (Pylyshyn and Storm, 1988), it operates as the following: An observer is presented with ca. 10 visually identical objects (dots, squares etc.) randomly distributed in the visual field. Various proportions of these objects are then identified as targets. Then all objects, which are visually indistinguishable, move around randomly and the subject's task is to track the target objects without moving his/her eyes. After a few seconds, the movement comes to a halt, one of the visual objects is marked, and the observer is asked to identify it as a target or non-target. One of the main findings of the MOT experiment is that a majority of observers is easily able to distinguish four or five of the randomly moving objects with

high accuracy (Pylyshyn and Storm, 1988). Furthermore, the number of objects to track does influence the observer's performance in accuracy and reaction time. Performance decreases monotonously with an increase in target objects.

Various theories and models have attempted to explain this phenomenon. The original authors (Pylyshyn and Storm, 1988) determined that their results were incompatible with the moving spot light theory traditionally described in the attentional shift literature, where a sole attention focus constantly rotates across all target positions in order to update them (Eriksen and Murphy, 1987; Posner, 1987). Pylyshyn instead established a theory that combines parallel and serial mechanisms in order to achieve the independent tracking of multiple visual objects in time and space (Pylyshyn and Storm, 1988). According to this visual indexing theory, each relevant object in the visual field receives a "finger of instantiation" (FINST). These are allotted pre-attentive in parallel and contain no further information about the object. Instead, they can be described as identifications without further contextual information. A tracking task can be described as follows: each relevant object on the screen is assigned one finger of the observer's hand (consequently, Pylyshyn described the maximum of track-able objects as 4-5), which attaches itself to one of the objects. Then each object's status is verified through serial scanning. Such serial scanning in combination with the upper limit of available FINST could explain the monotonously declining object tracking performance. Therefore, the problem is not the allocation of objects, but access to information (Pylyshyn, 1989). However, other reports have shown that the effortless maintenance of FINST is incompatible with several study results. So does increasing the tracking duration without feedback or an increase of set size result in an increase in errors where participants lost track of objects or confused them with distractors (Wolfe et al., 2007; Oksama and Hyönä, 2004). Furthermore, tracking performance declines when objects move at greater speed (Wolfe et al., 2007, Oksama and Hyönä, 2009) or are in close proximity to each other (Intriligator and Cavanagh, 2001; Alvarez and Franconeri, 2007). Allen et al. (2004) as well as Trick et al. (2006) also showed that the addition of a parallel, attention-demanding task lead

to a deterioration of individuals' tracking performance. Some have argued that MOT underlies a limited capacity processor (d'Avossa et al., 2006). However, a caveat of this theory is the limit in speed of attentional switching. Furthermore, studies have found independent resources for each visual hemifield, a bilateral advantage. Multiple object tracking performance of four visual objects in the lower visual field is comparable with that of two objects in the right or left hemifield (Alvarez and Cavanagh, 2005). Taken together, these findings rule out a single attention focus, and suggest that at least two independent attentional "spot lights" are present (Cavanagh and Alvarez, 2005).

An alternative explanatory approach is based on a resource-dependent process. Multiple spotlights, one for each attended object, share one global attention resource. Tracking performance is therefore defined by the task, such that when allocations of resources per object decrease, for example through increased speed or greater interferences, the ability to track each object deteriorates (Alvarez and Cavanagh, 2005; Cavanagh and Alvarez, 2005). Each target object receives an allocated portion of attentional resources, a FLEX (Alvarez and Franconeri, 2007) through the distribution of attentional resources. A series of reports supported this theory by showing that the neuronal networks activated during tracking coincide with structures of attention control (Culham et al., 1998; Howe et al., 2009). The determining factor of MOT is the resolution of each object's spotlight. This was proposed to be inversely correlated with the number of objects to track. The greater the amount of objects to track, the higher the likelihood over time to lose track of the target objects by confusing them with distractors (Intriligator and Cavanagh, 2001; Alvarez and Franconeri, 2007). In this context, the maintenance of object-relevant information is described as a function of visual working memory. Neuroimaging studies have found posterior parietal areas to exhibit attentional load effects and suggest this area as major force in the deployment of visual attention resources (Jovicich et al., 2001; Culham et al., 2001; Blumberg et al., 2015; Alnæs et al., 2017). Their results further confirmed the specific role of the intraparietal sulcus (IPS) in visual selective attention to provide feedback to visual areas in order to differentiate between task-relevant and task-

irrelevant responses to stimuli and are thus load dependent (Jovicich et al., 2001; Blumberg et al., 2015; Alnæs et al., 2017). Electrophysiological studies have also found an increase in event related potentials (ERPs) in relation to increased task loads in MOT tasks, specifically the contralateral delay activity (CDA), which manifests itself as a sustained negative wave over posterior contralateral electrode sites and hypothesized to stem from a source in the lateral IPS (Drew and Vogel, 2008; Drew et al., 2012; Luria et al., 2016) and is congruent with the visual working memory literature (Melcher and Piazza, 2011; Mazza and Caramazza, 2011). Additionally, velocity, number of distractors, and object proximity are further properties that can have interfering effects on MOT performance. The distribution of FLEXs across targets depends on the amount of interference imposed on each object, and therefore does not perforce follow an even distribution across targets (Bettencourt and Somers, 2009; Iordanescu et al., 2009; Clair et al., 2010; Franconeri et al., 2010; Franconeri et al., 2013; Drew et al., 2013). However, some studies report results that challenge the multifocal theory. For example, Holcomb and colleagues (2014) found that when the objects' velocity was high, performance increased when targets and distractors were closer to each other. They showed that object interference does not inevitably decrease tracking performance. Furthermore, Howe et al. (2010) found that tracking performance decreased in comparison to the classic MOT task (all target objects are moving) when only a fraction of targets were moving while the rest of the targets remained stationary. This result is incompatible with the multiple spotlight theory, as such a situation should lead to an increased proportion of resources or FLEXs for the moving target objects on one hand, while on the other hand, stationary target objects should require fewer resources/FLEXs.

Another approach is to consider study reports in which object tracking does not rely on the identities of each individual target, but instead considers the targets as an entity (Yantis, 1992; Jovicich et al., 2001; Merkel et al., 2015). For example Jovicich and colleagues reported that participants facilitated MOT by tracking imaginary 2D polygons formed by the target objects, which

is likely reflected in linear effects, increasing BOLD activation with evolving shape complexity, found in the "kinetic occipital" brain region, visual motion responsive areas in and posterior to V5/MT+ in the occipital cortex (Jovicich et al., 2001). Multivariate pattern analysis (MVPA) confirmed such findings (Merkel et al., 2015). Along those lines, but different in approach is the idea of using target merging in a MOT task in order to investigate the nature of visual object tracking. To this end, Scholl et al. (2001) used various procedures such as connecting lines or outlining borders/circumscription to unify one target with one distractor. Object tracking became more difficult to varying degrees depending on the merging technique, and in some cases subjects were only able to accurately track one object (Scholl et al., 2001). Therefore, certain combinations of objects, in which the combination itself becomes the focus of object-based attention, make it extremely difficult for an observer to differentiate the identity of individual objects.

However grouping of elements is not only facilitated through direct stimulus cues. Suganuma and Yokosawa (2006) observed declines in performance when targets and distractor objects exhibited similar movement trajectories, even when such pattern was not noticeable to participants. These results suggest that besides attentional external cues that draw attention to items, motion itself might also serve as a cue to form more global object representations. This finding could explain that the underlying cause of the impaired MOT performance observed in studies where only a portion of target elements moved (Howe et al., 2010) might be due to repression of grouping target elements into common movement trajectories. Experimental variations of the MOT task have revealed that multiple object tracking is not solely based on individual target information, nor a discrete reference principle originally described as FINSTs by Pylyshyn (1989). Even though participants correctly appropriate individual items to targets or distractors, they are unable to report each item's precise identity (Pylyshyn, 2004). Conclusively, it becomes apparent that neither of the above-mentioned theories is compatible with the multifaceted experimental MOT studies.

An alternative theory, albeit one that was published in the early days of MOT experiments but received little attention until recently, is by Yantis (Yantis, 1992). He demonstrated by manipulating the ease with which grouping was performed, that perceptual grouping of target elements, achieved either through specific movement trajectories or explicit instructions, improved MOT performance. "These findings support object-based theory of attention and demonstrate that perceptual grouping is not a purely stimulus-driven process, but can be governed by goal-directed mechanisms" (Yantis, 1992). This model describes MOT as a process that relies partially on the processing of global object information, that is, information that pertains to all target items. This is supported by empirical findings, for example, lack of individuation of targets along with temporal-spatial interactions, as described earlier. This global information processing of target items operates independently from the processing of individual target items, which is limited in resources and underlies capacity limitation (Miller, 1959; Luck and Vogel, 1997; Cowan, 2001).

In such ways, difficulties caused by manipulations in experimental factors, such as number of targets, motion speed or object spacing, can be compensated for (Yantis, 1992). Thus, it is plausible that MOT underlies two independent tracking processes: on the one hand, a global process whose focus is the abstract representation of a set of targets, while on the other hand, a local process that depends on the features of the individual targets. Work from Liu et al. (2005) has provided partial evidence for such independent processes. Participants were instructed to track multiple objects in a 3D scene that could undergo different viewpoint manipulations, which consequently altered the local movement properties of the objects. Interestingly this did not influence MOT performance. In respect thereof, MOT appears to rely on a higher-level representation of the whole scene with which local changes do not interfere. Albeit with an increase in target quantity, performance declined. In fact, studies mentioned earlier provided further evidence of a neurophysiological connection between target quantity and visual working memory.

Conclusively when investigating the MOT paradigm, it is crucial to differentiate between serial and parallel selection processes. In this context, the serial processing avenue refers not only to the temporal selection process (rotating spotlight) but also to the distribution of limited processing capacities to target items (multifocal theory). The main point in each of these theories is that an increase in task demands results in a decline in performance. In contrast, the parallel processing approach describes a segmentation of all target items into a higher-order object representation (Yantis, 1992). The current literature does not, or only partially, investigates the differences between these selection processes. Recent work by Merkel and colleagues (Merkel et al., 2014, 2015, 2017) has confirmed that participants utilize different strategies during an MOT task. In their experimental variation of the MOT task, they presented participants with a probe at the end of the task that highlighted a subset of the total items and could either be fully congruent, partially congruent or incongruent in terms of item identity location. This experimental design enabled them to disentangle participants' tracking strategies. By means of combined EEG and MEG recordings, they found that half of their participants utilized the global object tracking avenue which gave them behavioral advantages (faster reaction times and lower error rates) in trials in which the probe fully matched the tracked stimuli (Merkel et al., 2014). This was associated with enhanced probe-evoked neural activity at 170-210 ms (N180 component), relative to partial-match and no-match probes, in the lateral occipital cortex (Brodmann Area 19), a brain region associated with perceptual encoding of shapes and objects. The other half of the participants, who showed slower reaction times and higher error rates with increasing target probe correspondence, did not exhibit the N180 component effect. This provides evidence that these participants did not utilize the global tracking avenue, which was also reflected behaviorally. The researchers additionally found a N290 component at 270-310 ms, localized in the superior parietal cortex (BA 7) after the probe onset. This component was associated with the spatio-temporal properties of individual target items (mismatch effect), likely reflecting working memory processes and present in all participants (Merkel et al., 2014). Their results thus

provided evidence that MOT can be carried out by the two separate perceptual-cognitive strategies described here earlier. In a follow-up study, Merkel et al. (2017) investigated specifically the global object tracking strategy in which trained subjects tracked an imaginary shape constructed by the target items in a MOT task. Here they used a task-irrelevant luminance probe that could appear 1000-2000 ms after motion onset either outside, atop, or inside the putative shape created by the target items, and analyzed its resulting ERPs. Their findings suggested that attentional selection follows two time stages: Within 100-130 ms from the task-irrelevant probe, ERP responses to the probe atop the hypothetical figure differed from ERP responses to probes outside and inside the figure and were source-localized to the left lateral occipital cortex. This suggests that attention was first deployed to the boundaries of the global figure. In the succeeding 80 ms, ERP amplitude differed between outside and atop, as well as inside probes (while the latter two no longer differed from each other) and was source-localized within left ventral-occipital areas. These findings suggest that at this point, attention had spread to the full figure (Merkel et al., 2017).

Rhythmic processes in attention

While these results give some insight into the temporal mechanisms of object-based attention when subjects utilize global tracking mechanisms, it fails to shed light on the temporal differences between global and individual object processing.

Accumulating evidence suggests that rhythmic processes are functionally relevant for the covert allocation of selective attention (VanRullen et al., 2005, 2007). Rhythms in attention over time are typically explained as the brain's way of optimizing the so-called "exploration/exploitation dilemma" – the need to accumulate reliable information from each object while simultaneously monitoring other potentially relevant locations in the environment. A corpus of recent studies has demonstrated the rhythmic nature of attention and perception directly in human behavior (Landau and Fries, 2012; Fiebelkorn et al., 2013; VanRullen, 2013; Wutz et al., 2016; Ronconi and Melcher,

2017). These studies used a novel experimental method in which first trial-by-trial fluctuations of attention were reset to a specific point in time, and then behavioral performance was probed at several densely sampled time intervals after this reset event. This approach provides means to detect fine-scale temporal patterns in the allocation of selective attention, because behavior is sampled with high temporal resolution. Rhythmic fluctuations in attention have been reported both when aligned to sudden, external events (i.e. a briefly flashed spatial cue) and to transient, internal processing shifts triggered by a saccade. Two distinct frequency bands have been identified for behavioral rhythms, operating either on time scales around 100-150 ms (alpha frequency, 7-10 Hz) or 200-300 ms (theta frequency, 3-5 Hz) between successive samples, and different sampling frequencies are hypothesized to sub-serve different functions for selective attention (Fiebelkorn et al., 2013; Buschman and Kastner, 2015; Fries, 2015; VanRullen, 2016). Theta-frequency rhythms are found when attention switches between different objects, between different temporal events and between different spatial locations, and its putative role may lie in the sequential exploration of the visual environment. On these grounds, theta frequency would make an ideal candidate for the underlying rhythm that serves the global object-based, capacity-unlimited state of attention described in previous paragraphs. In turn, we propose faster alpha frequency rhythms to be the underlying mechanism that serve local object-based, capacity-limited states of attention, as they may reflect the rhythmic nature of "sustained" attention, mapping the same object representation over time by periodically sampling its spatio-temporal coordinates. Thus, capacity limits in object processing, as detailed in MOT studies, could result from the limited bandwidth of discrete, periodic computations, reflected in behavioral and brain rhythms. The MOT task is an ideal method to investigate this theory, as attention has to be maintained through time and space. In this experimental paradigm, attention is an ongoing process to maintain a set of different spatial configurations of target items at each temporal instance. According to the temporal windows model, objects are encoded in rhythmic cycles, and different objects are disambiguated by the phase angle in each cycle. Consequently,

object capacity limits reflect limited oscillatory bandwidth for the number of individual objects represented in each sampling cycle.

Motivation for the current study

Data from multiple object studies allowed researchers to show that the average position of multiple objects, the centroid target location, can be tracked even under conditions of reduced or withdrawn attention. For example, when the set size is above capacity limit or when the distractor centroid is probed (Alvarez and Oliva, 2008; Alvarez and Oliva, 2009; Alvarez, 2011; Drew and Chubb, 2010). These results suggest that information about multiple objects can be represented at an abstract level. Meaning, while such information lacks local detail, it contains a precise statistical summary of the scene (computing summary statistics).

With the current study, we set out to investigate how object-sampling rhythms relate to individual and global (average) object processing. In particular, we were interested in seeing if capacity-limited and capacity-unlimited states of selective attention were reflected in different sampling rhythms. The bottleneck for capacity limits in MOT and similar tasks arises on the level of object individuation. Object individuation involves selecting features from a crowded scene, binding them into a unitary object representation and individuating this unit from the background and other items in the scene (Xu and Chun, 2009; Wutz and Melcher, 2014). Individuation is classically measured with change detection for one individual item within a set of target items (partial report). We hypothesized that the two different object-tracking tasks would show different frequency patterns. In the present study, we used the method of rhythmic sampling of behavioral oscillations. As we hypothesized relatively short rhythms (lower range frequencies) to be the underlying mechanisms, behavioral oscillations lend themselves as a suitable method to investigate selective attention in individual and global object processing.

Materials and Methods

Participants

31 volunteers participated in the experiment (17 female, mean age 26.5 years; 5.9 standard deviation, one left-handed). Participants received monetary reimbursement and gave written informed consent for their participation. Experimental procedures adhered to the declaration of Helsinki and had prior approval by the Ethical Committee of the University of Trento.

Task & stimuli

The experiment was run on a CRT screen at 60 Hz with a spatial resolution of 1280 x 1024 pixel. Participants were seated approximately 52.5 cm from the monitor in a dimly lit room. Stimuli consisted of 8 white dots (0.5 degrees of visual angle) on a uniformly medium grey background. All dots were randomly placed in a centered, squared red frame (12 x 12 degrees of visual angle).

Each trial began with a central fixation cross (0.5 degrees of visual angle) within the centered square for 500 ms (Figure 14), and subjects were instructed to maintain fixation throughout the entire trial. Then the stimuli dots were presented and 4 of the dots were highlighted as targets by means of a bright green circle around each for 1000 ms. The subject's task was to attend and track these 4 moving dots while ignoring the four distractor dots. The initial movement direction of all dots was randomized, and then followed a linear path with a constant velocity of 2 degrees visual angle/s. These specific experimental parameters were chosen to ensure that participants were able to covertly attend the dynamic motion sequences with minimal eye movements and to ensure that the targets moved within the display region for the duration of a trial without bouncing back from the edges (thus avoiding an unnecessary confound). During the object motion, participants were either overtly asked to track each target object individually (individual tracking; highlighted in blue in Figure 14.A) or to group the target objects into a nonrigid virtual object and track its centroid position (average tracking; highlighted in red in Figure 14.B). During the tracking, the display was briefly

flashed with a white background for 50 ms at a randomly jittered time after the motion onset (0.5-1 s). On each trial the time interval of the post-flash motion was drawn pseudo-randomly from balanced time bins in steps of 50 ms between 0-1 s (20 Hz sampling). Then, the motion sequence stopped, and all stimulus dots disappeared for a memory delay of 1000 ms.

Then, in individual object tracking trials, all dots but one randomly chosen target dot reappeared in their respective locations at movement termination (thus, three target dots and all four distractor dots were presented stationary on the screen). Subjects were required to indicate the exact location of that absent target dot at the motion end (partial report). Whereas in average object tracking trials, none of the four targets reappeared, thus only the four distractor dots were visible, and subjects were required to indicate the centroid location of the targets. Subjects were instructed to respond as precisely as possible via mouse button clicks by making an unspeeded response. Tasks were blocked, and each participant performed four blocks (of 80 trials per block) of each task. Each block lasted approximately 10 minutes. Participants performed one practice block with 15-20 trials of each task to ensure they understood task instructions and to become familiar with the experiment. The stimulus sequence is illustrated in Figure 14.

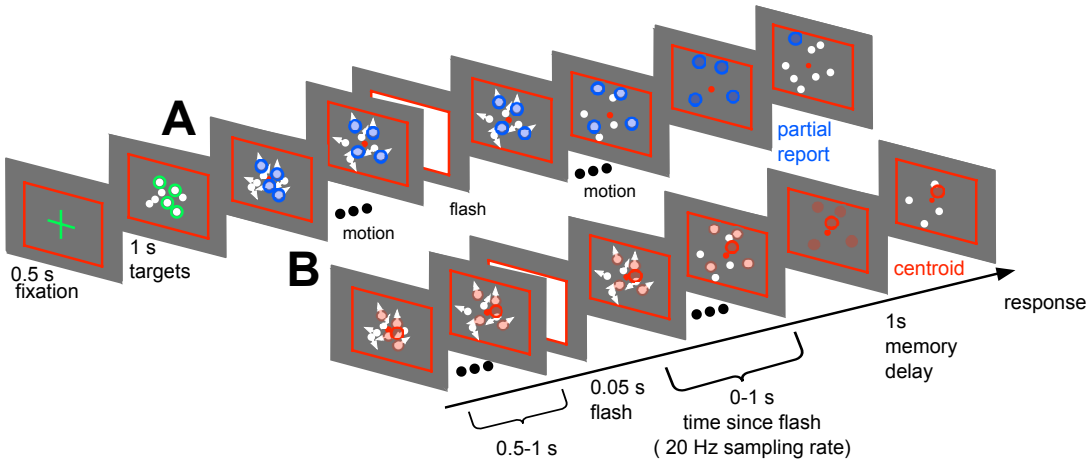

Figure 14. Schematic trial sequence of MOT task for both conditions. **A:** individual object tracking task. **B**: Average object tracking task. Target objects are highlighted in corresponding colors (individual in blue, average in red) for illustration purposes.

Data Analysis

Data were analyzed with MATLAB (MathWorks), and included functions from CircStat toolbox (Berens, 2009). Trials in which the object location error was below 1 degree of visual angle (v.a.) were considered as correct trials (proportion correct). Participants whose overall performance fell outside of three median absolute deviations from the median or reached ceiling effects at above 95 % accuracy in at least one condition were considered outliers and excluded from further analysis.

Then, trials were sorted into 20 equally spaced time bins (50-1000 ms in steps of 50 ms) by the temporal distance between the flash and the motion offset (Flash/Stimuli offset asynchrony, FSOA) and normalized within each subject separately. Specifically, in each subject, all trials were regarded as the time course distribution for that subject. Importantly, after the within-subject normalization, the relative relationship among the time course of all trials in each subject was kept intact, although the proportion-correct response values were normalized to 0. For each of the two conditions, the temporal profile of the normalized performance was calculated as a function of flash onset from 50 to 1000 ms in steps of 50 ms (20 Hz sampling frequency).

Our main goal was to examine the fine spectro-temporal dynamics of behavioral time courses. To do so, we analyzed the proportion-correct temporal courses within different frequencies through sinusoidal curve fitting and Fourier transformation and through the phase relationship between the different conditions.

Sinusoidal curve fitting

One assessment of the presence of behavioral oscillations is the fitting of a sinusoid to the behavioral fluctuations (Naitoh et al., 1985; Ronconi & Melcher, 2017). Data from both conditions were aggregated and averaged across participants with a smoothing factor of 20 Hz, resulting in a total of 301 trials for each of the 20 temporal bins for each condition. The individual object tracking and

average object-tracking time series were fitted separately with two independent sinusoidal functions with the following equation:

$$y = F(x) = \mu + \alpha \cos 2\pi f x + \beta \sin(2\pi f x)$$

Additionally, for each time series, we calculated the percent variance in the data that is explained by that sinus wave (goodness-of-fit, R^2 and adjusted R^2). All parameters were free except for the frequency (*f*), which was constrained to the frequency of interest (1-10 Hz).

Fourier transformation

To identify any periodic components relative to the flash onset in the time course of the data for each of the two conditions, the individual means of each subject were centered, and the data were Fourier-transformed with the application of a Hamming window and zero-padding (Fiebelkorn et al., 2011;Drewes et al., 2015; Ronconi & Melcher, 2017). Of the resulting Fourier spectra, the amplitude and phase information were averaged across subjects. A zero distribution was then generated by randomly shuffling the individual time points of the subjects (permutation analysis, N=10000), with subsequent (same procedure as before, Hamming window and zero padding) fast Fourier transformation (FFT). After sorting, the significance margin for the frequencies of interest (1-9 Hz) was determined by the percentage of zero distribution samples under the real averaged amplitude spectrum.

Finally, phase coherence constitutes support for oscillatory fluctuations in perception. As described above phase information (averaged across subjects) for each object-tracking time series was extracted from the Fourier spectra. Inter-trial coherence was calculated for each frequency of interest (1-9 Hz) separately for both tasks using the MATLAB CircStat toolbox (Berens, 2009). Additionally, the phase opposition product (POS; VanRullen, 2016b) between the two time series was calculated for each frequency bin of interest (1-9 Hz). A zero distribution was then generated

with the subsequent procedure as detailed above for amplitude spectrum in order to obtain significance boundaries for each frequency bin.

Results

Performance

Participants whose overall performance fell outside of three median absolute deviations from the median or reached ceiling effects at above 95 % accuracy in at least one condition were considered outliers and excluded from further analysis. Two subjects met our definition of outlier and were therefore excluded from all further analysis.

Average performance (N= 29) of individual object tracking ranged from 43.75 % to 83.13 % (mean =63.22 %, SD=10.47 %), whereas average performance of average object tracking ranged from 37.5 % to 89.69 % (mean= 68.53 %, SD=14.57 %). The difference was significant (paired two-tailed t-test, $t(28) = -2.5536$, $p= 0.0164$). Participants' error in individual object tracking was significantly higher than in average object tracking (error ± SD in visual angle; individual: 0.85° ± 0.19° v.a.; average: 0.6 ° ± 0.15° v.a.; $t(28) = 7.8519$, $p<0.001$). As Figure 15 illustrates, performance of the two tasks was correlated ($R= 0.64$, $p<0.001$). Participants that performed well in one task also performed well in the other task.

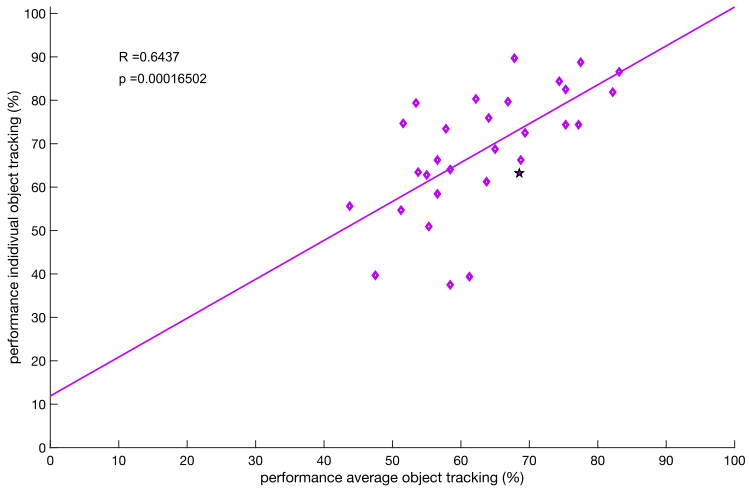

Figure 15. Task performance correlation. Each diamond depicts one subject's mean task performance in both tasks. Overall mean performance (N=29) marked by a star.

Curve fitting

The object tracking time courses relative to the onset of the flash revealed period fluctuations for both tasks in terms of percentage correct. On average, individual object tracking was best fit with a 6.63 Hz sinus function (explained variance, adjusted R^2 = 0.21; Figure 16.A), while average object tracking was best fit with a 4.57 Hz sinus function (explained variance, adjusted R^2 = 0.22; Figure16.B). Based on these results, the sinusoidal curve-fitting provides evidence for behavioral oscillations at different frequencies for the two object tracking tasks.

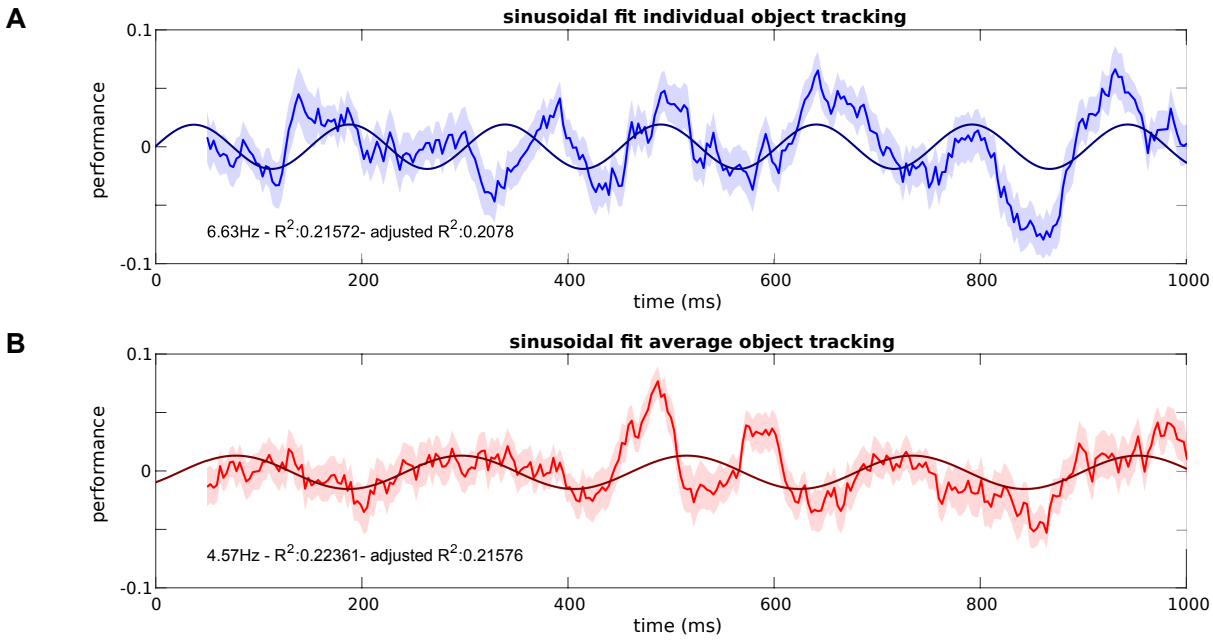

Figure 16. Mean accuracy time course (shaded areas show ± 1 SE) and best-fitting sinusoidal function in frequency range of interest (solid continuous line). **A:** individual object tracking task, **B**: average object tracking task. R^2 and adjusted R^2 values indicate goodness of fit for the respective sinusoidal function.

Fourier power analysis

For the individual object tracking task, we observed an average peak in the power spectrum at 8 Hz, with observed values significantly higher than the permutation spectrum ($p<0.05$; Figure 17.A).

While for the average object tracking results, we observed an average peak in the power spectrum at 5 Hz, with observed values showing a strong trend but failing to reach significance ($p>0.05$; Figure 17.B).

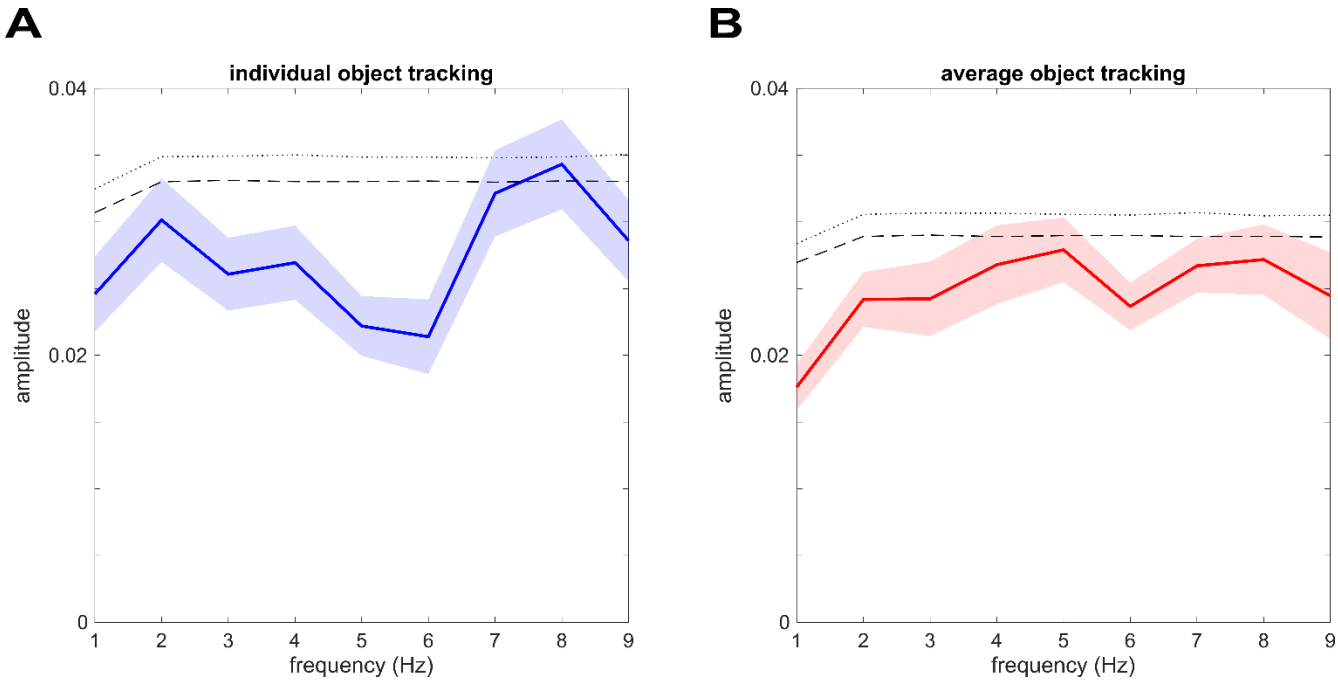

Figure 17. FFT power spectrum. FFT power Fourier analysis power spectrum of the behavioral performance for **A**: the individual object tracking task and **B**: the average object tracking task. Horizontal inserted lines in black show the 95% dashed) and 99% (dotted) significance margins obtained from the permutations test.

Fourier phase coherence analysis

One main peak at 8 Hz in the phase coherence spectrum of individual object tracking (Figure 18A) was found to be significant ($p<0.01$). One main peak at 5 Hz in the spectrum of average object tracking (Figure 18B) was found to show a strong trend towards significance ($p = 0.0518$).

Phase opposition (Figure 18C) was strongest for 3 Hz ($p= 0.0541$), as well as for 8 Hz ($p= 0.0672$). POS indicates strong phase coherence between participants but at polar-opposite phase angles. The two tasks alternated in counter phase in the 3 Hz behavioral rhythm (Figure 18D; mean difference phase angle = 149°).

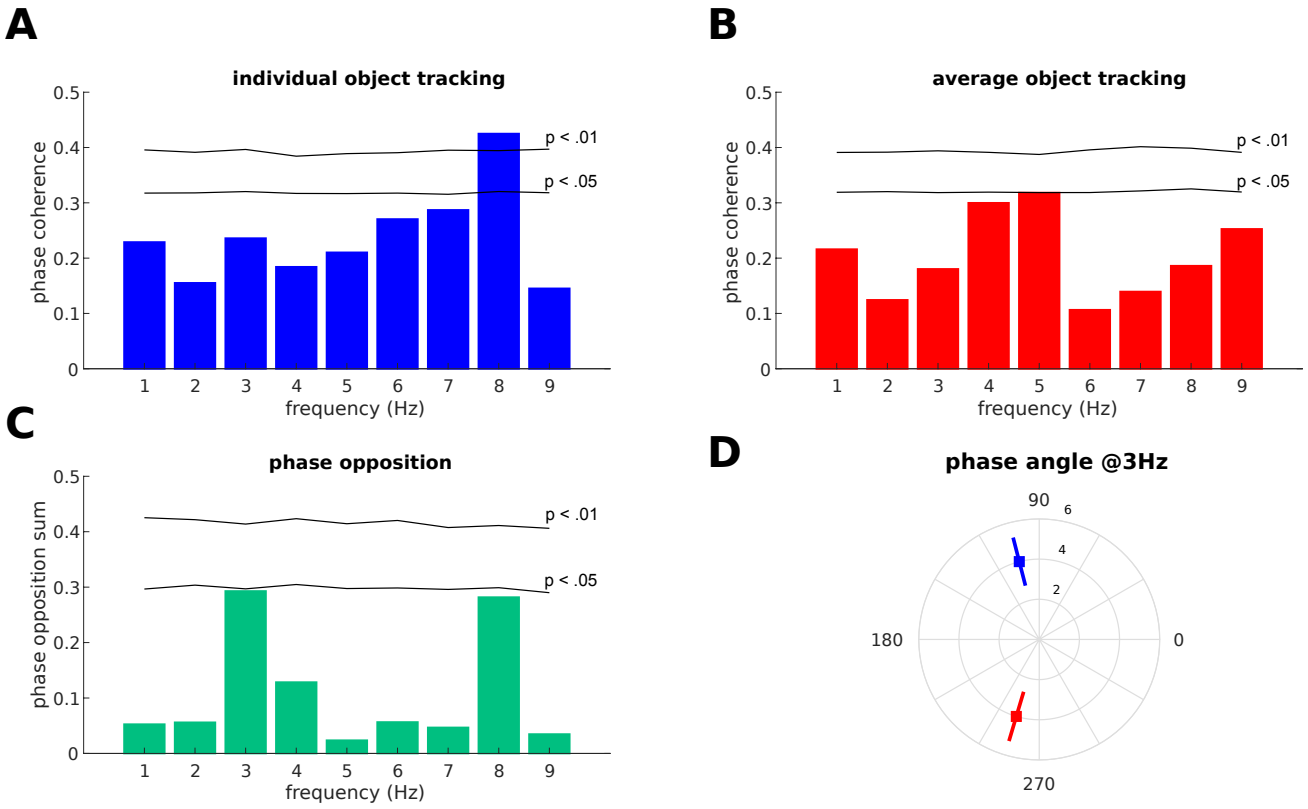

Figure 18. FFT phase results. **A**: phase coherence for individual object tracking task. **B**: phase coherence for average object tracking task. **C**: POS for both tasks. Black horizontal bars indicated permuted significance margins at 95 % and 99 %. **D**: mean phase angle histogram for 3 Hz for both behavioral oscillations (individual object tracking in blue, average object tracking in red, error bars depict circular standard deviation of phase angle at 3 Hz).

Discussion

As expected, participant's error rate in capacity-limited individual object tracking was significantly higher compared to average object tracking. Although task difficulty seemed comparable, participants performed better in the average object tracking task, which confirms previous MOT findings. Critically, the object tracking time course relative to the flash event revealed periodic fluctuations for both tasks in terms of percentage correct (Figure 16).

Attention seems to sample individual objects every 125-150 ms and to sample the average location between objects every 200-219 ms. These task-dependent behavioral rhythms were strongly consistent across participants as shown in the peaks in phase coherence in the frequency spectra of the task-accuracy time series (individual: 8 Hz, permutation test $p<0.01$; average: 5 Hz, $p= 0.0518$).

Moreover, the individual and average object memory time series were significantly shifted in phase at 3 Hz ($p<0.05$, Figure 18.C) as indicated by the phase opposition sum (VanRullen, 2016b). In fact, during tracking, average and individual object memory alternated in counter-phase in the behavioral 3 Hz rhythm, or, in other words, the individual location was represented 138 ms earlier than the average object location within the 333 ms object sampling cycle. The results of our study confirm that object properties are sampled rhythmically during multiple object tracking. In line with a temporal window account, we found a faster but low-bandwidth 7-8 Hz rhythm for individual objects and a slower but high-bandwidth 5 Hz rhythm for the object average. Thus, one possible explanation for capacity limitations might be based on discrete sampling cycles. Slower sampling could provide longer time windows for information integration from multiple objects, even for set sizes above the classic object-capacity limits, but in turn results in global pooling of local properties and a loss of precision for individual objects. In addition, the robust phase shift between individual and average object properties at the slower sampling rhythm might reflect a top-down strategy of the visual system to compensate for capacity limits. Local signals might be encoded first and then exist in alternation with global signals in order to coherently orchestrate the spatial representation of the scene layout and the objects contained therein.

Although subjects received explicit task instructions, it cannot be concluded that all participants actually utilized these particular tracking strategies. It is therefore possible that in the individual object tracking task, subjects tracked the objects as an imaginary global figure instead of each object individually. This is difficult to decipher in our current design. A modification of the task in which subjects are required to identify the probe object's identity (e.g. target number) could ensure individual tracking. In that sense, we cannot refute that our results confirm findings by Merkel et al. (2017) that the faster sampling rhythm of individual objects is representative of the imaginary shape building in which attention is directed to the outline composed of all targets, before attention spreads to the inside of the entire figure (which includes the centroid position) at later

latencies. In fact, the opposing phase angles at 3 Hz could be indicative of such temporal order of attentional deployment.

CHAPTER 5. GENERAL DISCUSSION

The main goal of the present thesis was to elucidate the role of oscillations in constructing multiple sensory objects in space and time. To this end, we conducted two MEG studies, which are reported in Chapter 2 and Chapter 3. We also conducted a behavioral oscillations study, which is reported in Chapter 4. In the following sections, I will briefly summarize these studies.

The first study was conducted to investigate the role of pre-stimulus alpha band oscillations in constructing visual objects over time. To achieve this, we utilized a paired-stimulus paradigm. Specifically, we presented participants with two brief flashes, where the first flash was at near-threshold perception level and varied the temporal gap between the two stimuli. After the stimulation, participants were to report the quantity of perceived flashes. In line with previous studies, longer temporal gaps resulted in increased performance. Furthermore, individual alpha peak frequency was predictive of perceptual outcome. Additionally, we found rhythmic fluctuations within the pre-stimulus alpha band by means of instantaneous frequency. This finding suggests that the alpha rhythm sped up slightly when participants correctly perceived the stimulus quantity.

In the second study, we investigated the role of neural oscillations in multisensory integration. To this end, we employed a spatial detection task. In particular, we designed a study in which we could manipulate multisensory attention, while keeping the stimulation constant but only changing the task. We presented temporally synchronous stimuli in the auditory, visual and somatosensory modality to participants. Participants were cued to attend to a specific stimulus pair and report whether it was spatially congruent or not. We found that both attention and stimulus congruence were reflected in behavioral benefits. Furthermore, we found pre-stimulus attention differences in the alpha and beta bands. In addition, we found attention effects in early evoked

responses at around 100 ms, which are suggestive of modality-specific integration differences in sensory regions.

The third study focused on the role of attention in sensory processing across time and space in a dynamic scene. In order to investigate the differences between local and global object attention, we utilized a variation of the multiple object tracking paradigm. Participants were presented with a set of identical moving objects amongst matching distractor objects. Their task was either to track the location of each discrete target item, or to track the targets items as a group. At specific times during the object motion, the screen was flashed with the intent of creating a phase reset in the visual system. The time course between the flash and the end of the object motion revealed strong oscillations in behavioral performance. Interestingly, the two tasks revealed different underlying perceptual cycles.

Overall, these studies provide support for the important role of neural oscillations, in particular the alpha band, in sensory processing. These findings are consistent with theories emphasizing the rhythmic nature of perception. In the case of pre-stimulus oscillations, these findings are consistent with the idea that brain rhythms are involved in the allocation of attention and in the temporal organization of continuous sensory input into discrete objects and events.

Perceptual and cognitive processes are limited by discrete, periodic computations in the brain, reflected in neural oscillations. It is impossible to simultaneously pay attention to everything in the dynamic environment that surrounds us. Human perception by nature is multimodal. Understanding how stimuli are encoded by different senses and how this information is combined in the most efficient way is also beneficial for applied domains such as robotics and virtual reality. Insights into the stimulus-perception relationship can allow for bypassing some technical limitations in designing virtual realities and robotics. This is already implemented in contemporary mobile phones, which are designed not only to ring, but also to vibrate and flash in order to alert the user efficiently through all

senses of an incoming call or message. A deeper understanding of neural oscillations and their dynamic interplay could be tremendously helpful in the improving the design of brain-computer interface technologies, such as robotic arms, neuroprostheses and therapeutic biofeedback devices that use brain activity recorded from the scalp as input (McFarland et al., 2010; Deuel et al., 2017). The combination of signals from multiple different cortical areas could improve the performance of neuroprostheses and other therapeutic biofeedback devices. It could ultimately enable individuals with motor disabilities to become more independent in navigating their environment with signals recorded from their scalp using non-invasive EEG electrodes. Brain wave monitoring devices could also monitor drivers' or pilots' mental states and automatically alert them if they fall asleep or become inattentive, and through brain-computer interface technology offer assistance in operating the vehicle or aircraft in challenging situations (Brown et al., 2013; Borghini et al., 2014).

Furthermore, a deeper understanding of sensory perception could have therapeutic value in a variety of pathologies that manifest attentional deficits or limited cognitive control, such as schizophrenia or attention deficit hyperactivity disorder. Understanding how the brain constructs (multiple) objects and events in space and time can be of use in the development of more focused treatment approaches for example through brain stimulation and pharmacological interventions that target and alter specific neural oscillations (Polanía et al., 2018).

Thus, understanding the underlying mechanisms that facilitate effective sensory processing and integration, their spatio-temporal profiles and determining what factors attention can have, is an important goal of neuroscience.

REFERENCES

Adrian, E. D., & Matthews, B. H. (1934). The Berger rhythm: potential changes from the occipital lobes in man. Brain, 57(4), 355-385.

Allen, R., Mcgeorge, P., Pearson, D., & Milne, A. B. (2004). Attention and expertise in multiple target tracking. Applied Cognitive Psychology, 18(3), 337-347.

Allison, T., McCarthy, G., Wood, C. C., Darcey, T. M., Spencer, D. D., & Williamson, P. D. (1989). Human cortical potentials evoked by stimulation of the median nerve. II. Cytoarchitectonic areas generating short-latency activity. Journal of neurophysiology, 62(3), 694-710.

Alnæs, D., Sneve, M. H., Richard, G., Skåtun, K. C., Kaufmann, T., Nordvik, J. E., Andreassen, O.A., Endestad, T., Laeng, B. & Westlye, L. T. (2015). Functional connectivity indicates differential roles for the intraparietal sulcus and the superior parietal lobule in multiple object tracking. Neuroimage, 123, 129-137.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. Current Biology, 15(9), 839-843.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. Trends in cognitive sciences, 15(3), 122-131.

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. Psychological science, 16(8), 637-643.

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? : Evidence for a resource-limited attentive tracking mechanism. Journal of vision, 7(13), 14-14.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. Psychological science, 19(4), 392-398.

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. Proceedings of the National Academy of Sciences, 106(18), 7345-7350.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annual review of neuroscience, 20(1), 303-330.

Azañón, E., & Soto-Faraco, S. (2008). Changing reference frames during the encoding of tactile events. Current Biology, 18(14), 1044-1049.

Badde, S., Röder, B., & Heed, T. (2014). Multiple spatial representations determine touch localization on the fingers. Journal of Experimental Psychology: Human Perception and Performance, 40(2), 784.

Başar-Eroglu, C., Strüber, D., Schürmann, M., Stadler, M., & Başar, E. (1996). Gamma-band responses in the brain: a short review of psychophysiological correlates and functional significance. International Journal of Psychophysiology, 24(1-2), 101-112.

Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (2001). Gamma, alpha, delta, and theta oscillations govern cognitive processes. International journal of psychophysiology, 39(2-3), 241-248.

Bauer, M., Oostenveld, R., Peeters, M., & Fries, P. (2006). Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital areas. Journal of Neuroscience, 26(2), 490-501.

Bauer, M., Kennett, S., & Driver, J. (2012). Attentional selection of location and modality in vision and touch modulates low-frequency activity in associated sensory cortices. Journal of neurophysiology, 107(9), 2342-2351.

Bauer, M., Stenner, M. P., Friston, K. J., & Dolan, R. J. (2014). Attentional modulation of alpha/beta and gamma oscillations reflect functionally distinct processes. Journal of Neuroscience, 34(48), 16117-16125.

Baumgarten, T. J., Schnitzler, A., & Lange, J. (2017). Beyond the Peak–Tactile Temporal Discrimination Does Not Correlate with Individual Peak Frequencies in Somatosensory Cortex. Frontiers in psychology, 8, 421.

Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. J Stat Softw, 31(10), 1-21.

Berger, H. (1929). Über das elektrenkephalogramm des menschen. Archiv für psychiatrie und nervenkrankheiten, 87(1), 527-570.

Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. Journal of vision, 9(7), 9-9.

Blanke O, Mohr C, Michel CM, Pascual-Leone A, Brugger P, Seeck M, Landis T, Thut G (2005) Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. J Neurosci 25:550–557, doi:10.1523/JNEUROSCI.2612-04.2005, pmid: 15659590.

Blumberg, E. J., Peterson, M. S., & Parasuraman, R. (2015). Enhancing multiple object tracking performance with noninvasive brain stimulation: a causal role for the anterior intraparietal sulcus. Frontiers in systems neuroscience, 9, 3.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neuroscience & Biobehavioral Reviews, 44, 58-75.

Buchholz, V. N., Jensen, O., & Medendorp, W. P. (2011). Multiple reference frames in cortical oscillatory activity during tactile remapping for saccades. Journal of Neuroscience, 31(46), 16864-16871.

Buneo, C. A., Batista, A. P., Jarvis, M. R., & Andersen, R. A. (2008). Time-invariant reference frames for parietal reach activity. Experimental brain research, 188(1), 77-89.

Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. Developmental psychobiology, 45(4), 204-220.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. Journal of Neuroscience, 29(24), 7869-7876.

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. Proceedings of the National Academy of Sciences, 107(37), 16048-16053.

Buschman, T. J., & Kastner, S. (2015). From behavior to neural dynamics: an integrated theory of attention. Neuron, 88(1), 127-144.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory–visual stimulus onset asynchrony detection. The Journal of Neuroscience, 21(1), 300-304.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. Science, 304(5679), 1926-1929.

Brainard, D. H. (1997). The psychophysics toolbox. Spatial vision, 10, 433-436.

Breitmeyer, B., & Öğmen, H. (2006). Visual masking: Time slices through conscious and unconscious vision (Vol. 41). Oxford University Press.

Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2008). Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events. Brain research bulletin, 75(6), 753-760.

Brown, T., Johnson, R., & Milavetz, G. (2013). Identifying periods of drowsy driving using EEG. Annals of advances in automotive medicine, 57, 99.

Calton, J. L., & Taube, J. S. (2009). Where am I and how will I get there from here? A role for posterior parietal cortex in the integration of spatial information and route planning. Neurobiology of learning and memory, 91(2), 186-196.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1493), 1001-1010.

Campisi, P., & La Rocca, D. (2014). Brain waves for automatic biometric-based user recognition. IEEE transactions on information forensics and security, 9(5), 782-800.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Current biology, 10(11), 649-657.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. Trends in cognitive sciences, 9(7), 349-354.

Cecere, R., Rees, G., & Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. Current Biology, 25(2), 231-235.

Clair, R. S., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. Journal of Vision, 10(4), 18-18.

Clark, C. R., Veltmeyer, M. D., Hamilton, R. J., Simms, E., Paul, R., Hermens, D., & Gordon, E. (2004). Spontaneous alpha peak frequency predicts working memory performance across the age span. International Journal of Psychophysiology, 53(1), 1-9.

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. Nature Reviews Neuroscience, 3(7), 553.

Cohen, M. X. (2014). Fluctuations in oscillation frequency control spike timing and coordinate neural networks. Journal of Neuroscience, 34(27), 8988-8998.

Cowan, N. (2001). Metatheory of storage capacity limits. Behavioral and brain sciences, 24(1), 154-176.

Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. Journal of neurophysiology, 80(5), 2657-2670.

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. Neuron, 32(4), 737-745.

d'Avossa, G., Shulman, G. L., Snyder, A. Z., & Corbetta, M. (2006). Attentional selection of moving objects by a serial process. Vision Research, 46(20), 3403-3412.

Debener, S., Herrmann, C. S., Kranczioch, C., Gembris, D., & Engel, A. K. (2003). Top-down attentional processing enhances auditory evoked gamma band activity. Neuroreport, 14(5), 683-686.

Deuel, T. A., Pampin, J., Sundstrom, J., & Darvas, F. (2017). The Encephalophone: A Novel Musical Biofeedback Device using Conscious Control of Electroencephalogram (EEG). Frontiers in human neuroscience, 11, 213.

Di Lollo, V. (1980). Temporal integration in visual memory. Journal of Experimental Psychology: General, 109(1), 75.

Di Pellegrino, G., & Làdavas, E. (2015). Peripersonal space in the brain. Neuropsychologia, 66, 126-133.

Dockree, P. M., Kelly, S. P., Roche, R. A., Hogan, M. J., Reilly, R. B., & Robertson, I. H. (2004). Behavioural and physiological impairments of sustained attention after traumatic brain injury. Cognitive Brain Research, 20(3), 403-414.

Donner, T. H., & Siegel, M. (2011). A framework for local cortical oscillation patterns. Trends in cognitive sciences, 15(5), 191-199.

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. Journal of Neuroscience, 28(16), 4183-4191.

Drew, S. A., Chubb, C. F., & Sperling, G. (2010). Precise attention filters for Weber contrast derived from centroid estimations. Journal of Vision, 10(10), 20-20.

Drew, T., Horowitz, T. S., Wolfe, J. M., & Vogel, E. K. (2012). Neural measures of dynamic changes in attentive tracking load. Journal of Cognitive Neuroscience, 24(2), 440-450.

Drew, T., Horowitz, T. S., & Vogel, E. K. (2013). Swapping or dropping? Electrophysiological measures of difficulty during multiple object tracking. Cognition, 126(2), 213-223.

Drewes, J., & VanRullen, R. (2011). This is the rhythm of your eyes: the phase of ongoing electroencephalogram oscillations modulates saccadic reaction time. Journal of Neuroscience, 31(12), 4698-4708.

Drewes, J., Zhu, W., Wutz, A., & Melcher, D. (2015). Dense sampling reveals behavioral oscillations in rapid visual categorization. Scientific reports, 5.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron, 57(1), 11-23.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations—signalling the status quo?. Current opinion in neurobiology, 20(2), 156-165.

Enns, J.T., & Di Lollo, V. (2000). What's new in visual masking? Trends in Cognitive Science, 4, 345-352.

Ergenoglu, T., Demiralp, T., Bayraktaroglu, Z., Ergen, M., Beydagi, H., & Uresin, Y. (2004). Alpha rhythm of the EEG modulates visual detection performance in humans. Cognitive Brain Research, 20(3), 376-383.

Fairhall, S. L., Albi, A., & Melcher, D. (2014). Temporal Integration Windows for Naturalistic Visual Sequences. PloS one, 9(7), e102248.

Felch, D. L., Khakhalin, A. S., & Aizenman, C. D. (2016). Multisensory integration in the developing tectum is constrained by the balance of excitation and inhibition. eLife, 5, e15600.

Fell, J., Fernandez, G., Klaver, P., Elger, C. E., & Fries, P. (2003). Is synchronized neuronal gamma activity relevant for selective attention?. Brain Research Reviews, 42(3), 265-272.

Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic sampling within and between objects despite sustained attention at a cued location. Current Biology, 23(24), 2553-2558.

Fontolan, L., Morillon, B., Liegeois-Chauvel, C., & Giraud, A. L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nature communications, 5, 4694.

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. Frontiers in psychology, 2, 154.

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Cognitive Brain Research, 10(1-2), 77-83.

Fracasso, A., Caramazza, A., & Melcher, D. (2010). Continuous perception of motion and shape across saccadic eye movements. Journal of Vision, 10, 14.

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. Psychological science, 21(7), 920-925.

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: competitive content maps for attention and memory. Trends in cognitive sciences, 17(3), 134-141.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. Experimental Brain Research, 147(3), 332-343.

Frey, J. N., Mainy, N., Lachaux, J. P., Müller, N., Bertrand, O., & Weisz, N. (2014). Selective modulation of auditory cortical alpha activity in an audiovisual spatial attention task. Journal of Neuroscience, 34(19), 6634-6639.

Frey, J. N., Ruhnau, P., & Weisz, N. (2015). Not so different after all: the same oscillatory processes support different types of attention. Brain research, 1626, 183-197.

Frey, J. N., Ruhnau, P., Leske, S., Siegel, M., Braun, C., & Weisz, N. (2016). The tactile window to consciousness is characterized by frequency-specific integration and segregation of the primary somatosensory cortex. Scientific reports, 6, 20805.

Fries, P. (2009). The model-and the data-gamma. Neuron, 64(5), 601-602.

Fries, P. (2015). Rhythms for cognition: communication through coherence. Neuron, 88(1), 220-235.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. Science, 291(5508), 1560-1563.

Fu, K. M. G., Foxe, J. J., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2001). Attention-dependent suppression of distracter visual input can be cross-modally cued as indexed by anticipatory parieto–occipital alpha-band oscillations. Cognitive Brain Research, 12(1), 145-152.

Fukuda, M., Nishida, M., Juhasz, C., Muzik, O., Sood, S., Chugani, H. T., & Asano, E. (2008). Short-latency median-nerve somatosensory-evoked potentials and induced gamma-oscillations in humans. Brain, 131(7), 1793-1805.

Ghazanfar, A. A., & Logothetis, N. K. (2003). Neuroperception: Facial expressions linked to monkey calls. Nature, 423(6943), 937-938.

Gips, B., Eerden, J. P., & Jensen, O. (2016). A biologically plausible mechanism for neuronal coding organized by the phase of alpha oscillations. European Journal of Neuroscience, 44(4), 2147-2161.

Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., Schwartz, T. H., & Foxe, J. J. (2011). Oscillatory sensory selection mechanisms during intersensory attention to rhythmic auditory and

visual inputs: a human electrocorticographic investigation. Journal of Neuroscience, 31(50), 18556-18567.

Gramann, K. (2013). Embodiment of spatial reference frames and individual differences in reference frame proclivity. Spatial Cognition & Computation, 13(1), 1-25.

Grandy, T. H., Werkle-Bergner, M., Chicherio, C., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2013). Individual alpha peak frequency is related to latent factors of general cognitive abilities. Neuroimage, 79, 10-18.

Gratton, G. (2018). Brain reflections: A circuit-based framework for understanding information processing and cognitive control. Psychophysiology, 55(3).

Graziano, M. S., Hu, X. T., & Gross, C. G. (1997). Visuospatial properties of ventral premotor cortex. Journal of Neurophysiology, 77(5), 2268-2292.

Graziano, M. S., Reiss, L. A., & Gross, C. G. (1999). A neuronal representation of the location of nearby sounds. Nature, 397(6718), 428-430.

Gregoriou, G. G., Paneri, S., & Sapountzis, P. (2015). Oscillatory synchrony as a mechanism of attentional processing. Brain research, 1626, 165-182.

Hadjidimitrakis, K., Bertozzi, F., Breveglieri, R., Galletti, C., & Fattori, P. (2017). Temporal stability of reference frames in monkey area V6A during a reaching task in 3D space. Brain Structure and Function, 222(4), 1959-1970.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. The Journal of Neuroscience, 28(10), 2539-2550.

Heed, T., Buchholz, V. N., Engel, A. K., & Röder, B. (2015). Tactile remapping: from coordinate transformation to integration in sensorimotor processing. Trends in cognitive sciences, 19(5), 251-258.

Herrmann, C. S., Munk, M. H., & Engel, A. K. (2004). Cognitive functions of gamma-band activity: memory match and utilization. Trends in cognitive sciences, 8(8), 347-355.

Hlushchuk, Y., & Hari, R. (2006). Transient suppression of ipsilateral primary somatosensory cortex during tactile finger stimulation. The Journal of neuroscience, 26(21), 5819-5824.

Holcombe, A. O., Chen, W. Y., & Howe, P. D. (2014). Object tracking: Absence of long-range spatial interference supports resource theories. Journal of vision, 14(6), 1-1.

Howe, P. D., Horowitz, T. S., Morocz, I. A., Wolfe, J., & Livingstone, M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. Journal of Vision, 9(4), 10-10.

Howe, P. D., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. Journal of Vision, 10(8), 11-11.

Iemi, L., Chaumon, M., Crouzet, S. M., & Busch, N. A. (2017). Spontaneous neural oscillations bias perception by modulating baseline excitability. Journal of Neuroscience, 37(4), 807-819.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. Cognitive psychology, 43(3), 171-216.

Ioannides, A. A., Liu, L., Poghosyan, V., Saridis, G. A., Gjedde, A., Ptito, M., & Kupers, R. (2013). MEG reveals a fast pathway from somatosensory cortex to occipital areas via posterior parietal cortex in a blind subject. Frontiers in Human Neuroscience, 7, 429.

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. Journal of vision, 9(4), 1-1.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the" ventriloquism" effect. Perceptual and motor skills.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Frontiers in human neuroscience, 4, 186.

Jensen, O., Kaiser, J., & Lachaux, J. P. (2007). Human gamma-frequency oscillations associated with attention and memory. Trends in neurosciences, 30(7), 317-324.

Jiang, Y., Kumar, A., & Vickery, T. J. (2005). Integrating sequential arrays in visual short-term memory. Experimental Psychology, 52(1), 39.

Jin, Y., O'HALLORAN, J. P., Plon, L., Sandman, C. A., & Potkin, S. G. (2006). Alpha EEG predicts visual reaction time. International Journal of Neuroscience, 116(9), 1035-1044.

Jones EG, Powell TP (1970) An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain 93:793–820, doi:10.1093/brain/93.4.793, pmid: 4992433.

Jones, S. R., Kerr, C. E., Wan, Q., Pritchett, D. L., Hämäläinen, M., & Moore, C. I. (2010). Cued spatial attention drives functionally relevant modulation of the mu rhythm in primary somatosensory cortex. Journal of Neuroscience, 30(41), 13760-13765.

Jovicich, J., Peters, R. J., Koch, C., Braun, J., Chang, L., & Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. Journal of cognitive neuroscience, 13(8), 1048-1058.

Kaulmann, D., Hermsdörfer, J., & Johannsen, L. (2017). Disruption of right posterior parietal cortex by continuous Theta Burst Stimulation alters the control of body balance in quiet stance. European Journal of Neuroscience, 45(5), 671-678.

Karns CM, Dow MW, Neville HJ. (2012) Altered Cross-Modal Processing in the Primary Auditory Cortex of Congenitally Deaf Adults: A Visual-Somatosensory fMRI Study with a Double-Flash Illusion. The Journal of Neuroscience 32(28): 9626-9638.

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. Neuron, 48(2), 373-384.

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. Cerebral Cortex, 18(7), 1560-1574.

Keil, J., Pomper, U., Feuerbach, N., & Senkowski, D. (2017). Temporal orienting precedes intersensory attention and has opposing effects on early evoked brain activity. NeuroImage, 148, 230-239.

Kelly, S. P., Lalor, E. C., Reilly, R. B., & Foxe, J. J. (2006). Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. Journal of neurophysiology, 95(6), 3844-3851.

Kendall, M. G., & Buckland, W. R. (1971). A dictionary of statistical terms. New York: Hafner.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain research reviews, 29(2-3), 169-195.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. Trends in cognitive sciences, 16(12), 606-617.

Klimesch, W., Schimke, H., & Pfurtscheller, G. (1993). Alpha frequency, cognitive load and memory performance. Brain topography, 5(3), 241-251.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition–timing hypothesis. Brain research reviews, 53(1), 63-88.

Kida, T., Inui, K., Wasaka, T., Akatsuka, K., Tanaka, E., & Kakigi, R. (2007). Time-varying cortical activations related to visual–tactile cross-modal links in spatial selective attention. Journal of neurophysiology, 97(5), 3585-3596.

Kondacs, A., & Szabó, M. (1999). Long-term intra-individual variability of the background EEG in normals. Clinical Neurophysiology, 110(10), 1708-1716.

Lacquaniti, F., & Viviani, P. (Eds.). (1996). Neural bases of motor behaviour (Vol. 85). Springer Science & Business Media.

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. Current biology, 22(11), 1000-1004.

Lange, J., Oostenveld, R., & Fries, P. (2013). Reduced occipital alpha power indexes enhanced excitability rather than improved visual perception. Journal of Neuroscience, 33(7), 3212-3220.

Lange, J., Keil, J., Schnitzler, A., van Dijk, H., & Weisz, N. (2014). The role of alpha oscillations for illusory perception. Behavioural brain research, 271, 294-301.

Lansbergen, M. M., Arns, M., van Dongen-Boomsma, M., Spronk, D., & Buitelaar, J. K. (2011). The increase in theta/beta ratio on resting-state EEG in boys with attention-deficit/hyperactivity disorder is mediated by slow alpha peak frequency. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 35(1), 47-52.

Lemus, L., Hernández, A., Luna, R., Zainos, A., & Romo, R. (2010). Do sensory cortices process more than one sensory modality during perceptual judgments?. Neuron, 67(2), 335-348.

Leonardelli, E., Braun, C., Weisz, N., Lithari, C., Occelli, V., & Zampini, M. (2015). Prestimulus oscillatory alpha power and connectivity patterns predispose perceptual integration of an audio and a tactile stimulus. Human brain mapping, 36(9), 3486-3498.

Leske, S., Ruhnau, P., Frey, J., Lithari, C., Müller, N., Hartmann, T., & Weisz, N. (2015). Prestimulus network integration of auditory cortex predisposes near-threshold perception independently of local excitability. Cerebral Cortex, 25(12), 4898-4907.

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., & Enns, J. T. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. Journal of Experimental Psychology: Human Perception and Performance, 31(2), 235.

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. Neuroscience & Biobehavioral Reviews, 62, 100-108.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. Nature, 390(6657), 279.

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. Science, 289(5482), 1206-1208.

Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The curious incident of attention in multisensory integration: bottom-up vs. top-down. Multisensory Research, 29(6-7), 557-583.

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To see or not to see: prestimulus α phase predicts visual awareness. Journal of Neuroscience, 29(9), 2725-2732.

Mazaheri, A., van Schouwenburg, M. R., Dimitrijevic, A., Denys, D., Cools, R., & Jensen, O. (2014). Region-specific modulations in oscillatory alpha activity serve to facilitate processing in the visual and auditory modalities. Neuroimage, 87, 356-362.

Mazza, V., & Caramazza, A. (2011). Temporal brain dynamics of multiple object processing: the flexibility of individuation. PloS one, 6(2), e17453.

McDonald, J. J., Teder-Sälejärvi, W. A., Di Russo, F., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. Journal of Cognitive Neuroscience, 15(1), 10-19.

McDonald, J. J., & Ward, L. M. (2000). Involuntary listening aids seeing: Evidence from human electrophysiology. Psychological Science, 11(2), 167-171.

McFarland, D. J., Sarnacki, W. A., & Wolpaw, J. R. (2010). Electroencephalographic (EEG) control of three-dimensional movement. Journal of neural engineering, 7(3), 036007.

McGuire, L. M., & Sabes, P. N. (2011). Heterogeneous representations in the superior parietal lobule are common across reaches to visual and proprioceptive targets. Journal of Neuroscience, 31(18), 6661-6673.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

Melcher, D., & Piazza, M. (2011). The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. PloS one, 6(12), e29296.

Melcher, D., & Fracasso, A. (2012). Remapping of the line motion illusion across eye movements. Experimental Brain Research, 218, 503-514.

Melcher, D., Wutz, A., Drewes, J., & Fairhall, S. (2014). The role of temporal integration windows in visual perception. Procedia-Social and Behavioral Sciences, 126, 92-93.

Merkel, C., Stoppel, C. M., Hillyard, S. A., Heinze, H. J., Hopf, J. M., & Schoenfeld, M. A. (2014). Spatio-temporal patterns of brain activity distinguish strategies of multiple-object tracking. Journal of cognitive neuroscience, 26(1), 28-40.

Merkel, C., Hopf, J. M., Heinze, H. J., & Schoenfeld, M. A. (2015). Neural correlates of multiple object tracking strategies. Neuroimage, 118, 63-73.

Merkel, C., Hopf, J. M., & Schoenfeld, M. A. (2017). Spatio-temporal dynamics of attentional selection stages during multiple object tracking. NeuroImage, 146, 484-491.

Meredith, M. A., Allman, B. L., Keniston, L. P., & Clemo, H. R. (2012). Are bimodal neurons the same throughout the brain? In: Wallace MT, Murray MM (eds) Frontiers in the neural bases of multisensory processing.

Miller, G. A. (1956). GA Miller, Psychol. Rev. 63, 81 (1956). Psychol. Rev., 63, 81.

Moretti, D. V., Babiloni, C., Binetti, G., Cassetta, E., Dal Forno, G., Ferreric, F., ... & Rodriguez, G. (2004). Individual analysis of EEG frequency and band power in mild Alzheimer's disease. Clinical Neurophysiology, 115(2), 299-308.

Morrell, F. (1972). Visual system's view of acoustic space. Nature, 238(5358), 44.

Müller, N., & Weisz, N. (2011). Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. Cerebral Cortex, 22(7), 1604-1613.

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., Schroeder, C.E., & Foxe, J. J. (2005). Grabbing your ear: rapid auditory–somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cerebral Cortex, 15(7), 963-974.

Naitoh, P., Englund, C. E., & Ryman, D. H. (1985). Circadian rhythms determined by cosine curve fitting: Analysis of continuous work and sleep-loss data. Behavior Research Methods, Instruments, & Computers, 17(6), 630-641.

Noel, J. P., & Wallace, M. (2016). Relative contributions of visual and auditory spatial representations to tactile localization. Neuropsychologia, 82, 84-90.

Nunez, P. L., & Srinivasan, R. (2010). Scale and frequency chauvinism in brain dynamics: too much emphasis on gamma band oscillations. Brain Structure and Function, 215(2), 67-71.

Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. Visual cognition, 11(5), 631-671.

Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. Cognitive psychology, 56(4), 237-283.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational intelligence and neuroscience, 2011, 1.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. Frontiers in neuroinformatics, 10, 27.

Pasalar, S., Ro, T., & Beauchamp, M. S. (2010). TMS of posterior parietal cortex disrupts visual tactile multisensory integration. European Journal of Neuroscience, 31(10), 1783-1790.

Pavani, F., Spence, C., & Driver, J. (2000). Visual capture of touch: Out-of-the-body experiences with rubber gloves. Psychological science, 11(5), 353-359.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial vision, 10(4), 437-442.

Penfield, W., & Rasmussen, T. (1950). The cerebral cortex of man; a clinical study of localization of function.

Polanía, R., Nitsche, M. A., & Ruff, C. C. (2018). Studying and modifying brain function with non-invasive brain stimulation. Nature neuroscience, 1.

Pomper et al., 2014.Taking a Call is Facilitated by the multisensory processing of smartphone vibrations, sounds and flashes. PLOS ONE.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. Cognition, 32(1), 65-97.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. Spatial vision, 3(3), 179-197.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. Cognition, 80, 127–158.

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. Visual cognition, 11(7), 801-822.

Romei, V., Brodbeck, V., Michel, C., Amedi, A., Pascual-Leone, A., & Thut, G. (2008a). Spontaneous fluctuations in posterior α-band EEG activity reflect variability in excitability of human visual areas. Cerebral cortex, 18(9), 2010-2018.

Romei, V., Rihs, T., Brodbeck, V., & Thut, G. (2008b). Resting electroencephalogram alpha-power over posterior sites indexes baseline visual cortex excitability. Neuroreport, 19(2), 203-208.

Romei, V., Gross, J., & Thut, G. (2010). On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: correlation or causation?. Journal of Neuroscience, 30(25), 8692-8697.

Romei, V., Gross, J., & Thut, G. (2012). Sounds reset rhythms of visual cortex and corresponding human visual perception. Current biology, 22(9), 807-813.

Ronconi, L., & Marotti, R. B. (2017). Awareness in the crowd: Beta power and alpha phase of prestimulus oscillations predict object discrimination in visual crowding. Consciousness and cognition, 54, 36-46.

Ronconi, L., & Melcher, D. (2017). The role of oscillatory phase in determining the temporal organization of perception: Evidence from sensory entrainment. Journal of Neuroscience, 37(44), 10636-10644.

Ronconi, L., Oosterhof, N. N., Bonmassar, C., & Melcher, D. (2017). Multiple oscillatory rhythms determine the temporal organization of perception. Proceedings of the National Academy of Sciences, 114(51), 13435-13440.

Ruhnau, P., Hauswald, A., & Weisz, N. (2014). Investigating ongoing brain oscillations and their influence on conscious perception–network states and the window to consciousness. Frontiers in psychology, 5, 1230.

Samaha, J., & Postle, B. R. (2015). The speed of alpha-band oscillations predicts the temporal resolution of visual perception. Current Biology, 25(22), 2985-2990.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neuroscience letters, 127(1), 141-145.

Sarri, M., Blankenburg, F., & Driver, J. (2006). Neural correlates of crossmodal visual-tactile extinction and of tactile awareness revealed by fMRI in a right-hemisphere stroke patient. Neuropsychologia, 44(12), 2398-2410.

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. Current opinion in neurobiology, 20(2), 172-176.

Scholl, B. (2017, August 11). Reference Guide: Multiple Object Tracking. Retrieved February 13, 2018, from http://perception.yale.edu/Brian/refGuides/MOT.html

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. Cognition, 80(1-2), 159-177.

Schürmann, M., Caetano, G., Hlushchuk, Y., Jousmäki, V., & Hari, R. (2006). Touch activates human auditory cortex. Neuroimage, 30(4), 1325-1331.

Schubert, R., Haufe, S., Blankenburg, F., Villringer, A., & Curio, G. (2009). Now you'll feel it, now you won't: EEG rhythms predict the effectiveness of perceptual masking. Journal of cognitive neuroscience, 21(12), 2407-2419.

Sekiyama, K., &Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception.Journal of Phonetics, 21, 427–444.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. Perception & Psychophysics, 59(1), 73-80.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia, 45(3), 561-571.

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. Cognitive psychology, 43(4), 274-310.

Shore, D. I., Spry, E., & Spence, C. (2002). Confusing the mind by crossing the hands. Cognitive Brain Research, 14(1), 153-163.

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations?. Neuron, 24(1), 49-65.

Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. Perception & psychophysics, 65(2), 318-328.

Spence, C., & Driver, J. (2004). Crossmodal space and crossmodal attention. Oxford University Press.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. Current Biology, 13(13), R519-R521.

Sperdin, H. F., Cappe, C., Foxe, J. J., & Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. Frontiers in integrative neuroscience, 3, 2.

Stein, B. E., Stanford, T. R., Wallace, M. T., Vaughan, J. W., & Jiang, W. (2004). Crossmodal spatial interactions in subcortical and cortical circuits. Crossmodal space and crossmodal attention, 25-50.

Stein, J. F. (1989). Representation of egocentric space in the posterior parietal cortex. Quarterly Journal of Experimental Physiology, 74(5), 583-606.

Stein, J. F. (1992). The representation of egocentric space in the posterior parietal cortex. Behavioral and Brain Sciences, 15(04), 691-700.

Soto-Faraco, S., & Alsius, A. (2007). Conscious access to the unisensory components of a cross-modal illusion. Neuroreport, 18(4), 347-350.

Soto-Faraco, S., Ronald, A., & Spence, C. (2004). Tactile selective attention and body posture: assessing the multisensory contributions of vision and proprioception. Perception & Psychophysics, 66(7), 1077-1094.

Suganuma, M., & Yokosawa, K. (2006). Grouping and trajectory storage in multiple object tracking: Impairments due to common item motions. Perception, 35(4), 483-495.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. Trends in cognitive sciences, 3(4), 151-162.

Tallon-Baudry, C. (2003). Oscillatory synchrony and human visual cognition. Journal of Physiology-Paris, 97(2-3), 355-363.

Taulu, S., & Hari, R. (2009). Removal of magnetoencephalographic artifacts with temporal signal-space separation: demonstration with single-trial auditory-evoked responses. Human brain mapping, 30(5), 1524-1534.

Teder-Sälejärvi, W. A., Russo, F. D., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. Journal of cognitive neuroscience, 17(9), 1396-1409.

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). α-Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. Journal of Neuroscience, 26(37), 9494-9502.

Tiippana, K., Andersen, T. S., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. European Journal of Cognitive Psychology, 16(3), 457-472.

Tiitinen, H. T., Sinkkonen, J., Reinikainen, K., Alho, K., Lavikainen, J., & Näätänen, R. (1993). Selective attention enhances the auditory 40-Hz transient response in humans. Nature, 364(6432), 59.

Trick, L. M., Guindon, J., & Vallis, L. A. (2006). Short Article: Sequential Tapping Interferes Selectively with Multiple-Object Tracking: Do Finger-Tapping and Tracking Share a Common Resource?. Quarterly Journal of Experimental Psychology, 59(7), 1188-1195.

Van Dijk, H., Schoffelen, J. M., Oostenveld, R., & Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. Journal of Neuroscience, 28(8), 1816-1823.

Van Ede, F., de Lange, F., Jensen, O., & Maris, E. (2011). Orienting attention to an upcoming tactile event involves a spatially and temporally specific modulation of sensorimotor alpha-and beta-band oscillations. Journal of Neuroscience, 31(6), 2016-2024.

VanGilder, P., Shi, Y., Apker, G., Dyson, K., & Buneo, C. A. (2016). Multisensory Interactions Influence Neuronal Spike Train Dynamics in the Posterior Parietal Cortex. PloS one, 11(12), e0166786.

VanRullen, R. (2013). Visual attention: a rhythmic process?. Current Biology, 23(24), R1110-R1112.

VanRullen, R. (2016). Perceptual cycles. Trends in Cognitive Sciences, 20(10), 723-735.

VanRullen, R. (2016b). How to evaluate phase differences between trial groups in ongoing electrophysiological signals. Frontiers in neuroscience, 10, 426.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous?. Trends in cognitive sciences, 7(5), 207-213.

VanRullen, R., Reddy, L., & Koch, C. (2005). Attention-driven discrete sampling of motion perception. Proceedings of the National Academy of Sciences of the United States of America, 102(14), 5291-5296.

VanRullen, R., Reddy, L., & Koch, C. (2006). The continuous wagon wheel illusion is associated with changes in electroencephalogram power at~ 13 Hz. Journal of Neuroscience, 26(2), 502-507.

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. Proceedings of the National Academy of Sciences, 104(49), 19204-19209.

VanRullen, R., & Dubois, J. (2011). The psychophysics of brain rhythms. Frontiers in psychology, 2, 203.

VanRullen, R., Busch, N., Drewes, J., & Dubois, J. (2011). Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability. Frontiers in psychology, 2, 60.

VanRullen, R., Zoefel, B., & Ilhan, B. (2014). On the cyclic nature of perception in vision versus audition. Philosophical transactions of the Royal Society B: Biological sciences, 369(1641), 20130214.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. Neuropsychologia, 45(3), 598-607.

Van Wassenhove, V. (2013). Speech through ears and eyes: interfacing the senses with the supramodal brain. Frontiers in psychology, 4, 388.

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. Experimental Brain Research, 158(2), 252-258.

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. Physiological reviews, 90(3), 1195-1268.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. Perception & psychophysics, 33(2), 113-120.

Weisz, N., Wühle, A., Monittola, G., Demarchi, G., Frey, J., Popov, T., & Braun, C. (2014). Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. Proceedings of the National Academy of Sciences, 111(4), E417-E425.

Wolfe, J. M., Place, S. S., & Horowitz, T. S. (2007). Multiple object juggling: Changing what is tracked during extended multiple object tracking. Psychonomic bulletin & review, 14(2), 344-349.

Wutz, A., & Melcher, D. (2014). The temporal window of individuation limits visual capacity. Frontiers in psychology, 5, 952.

Wutz, A., Weisz, N., Braun, C., & Melcher, D. (2014). Temporal Windows in Visual Processing: "Prestimulus Brain State" and "Poststimulus Phase Reset" Segregate Visual Transients on Different

Temporal Scales. The Journal of Neuroscience 34(4), 1554-1565

Wutz, A., Muschter, E., van Koningsbruggen, M. G., Weisz, N., & Melcher, D. (2016). Temporal integration windows in neural processing and perception aligned to saccadic eye movements. Current Biology, 26(13), 1659-1668.

Wutz, A., Melcher, D., & Samaha, J. (2018). Frequency modulation of neural oscillations according to visual task demands. Proceedings of the National Academy of Sciences, 201713318.

Wyart, V., & Tallon-Baudry, C. (2009). How ongoing fluctuations in human visual cortex predict perceptual awareness: baseline shift versus decision bias. Journal of Neuroscience, 29(27), 8715-8725.

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. Trends in cognitive sciences, 13(4), 167-174.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. Cognitive psychology, 24(3), 295-340.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. Experimental brain research, 152(2), 198-210.

Zampini, M., Brown, T., Shore, D. I., Maravita, A., Röder, B., & Spence, C. (2005). Audiotactile temporal order judgments. Acta psychologica, 118(3), 277-291.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. Cerebral cortex, 11(10), 946-953.